

# Seeing in the dark – I. Multi-epoch alchemy

Eric M. Huff,<sup>1★</sup> Christopher M. Hirata,<sup>2</sup> Rachel Mandelbaum,<sup>3,4</sup> David Schlegel,<sup>5</sup>  
Uroš Seljak<sup>5,6,7,8</sup> and Robert H. Lupton<sup>3</sup>

<sup>1</sup>Department of Astronomy, University of California at Berkeley, Berkeley, CA 94720, USA

<sup>2</sup>Department of Astronomy, Caltech M/C 350-17, Pasadena, CA 91125, USA

<sup>3</sup>Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA

<sup>4</sup>Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

<sup>5</sup>Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

<sup>6</sup>Space Sciences Lab, Department of Physics and Department of Astronomy, University of California, Berkeley, CA 94720, USA

<sup>7</sup>Institute of the Early Universe, Ewha Womans University, Seoul 120-750, South Korea

<sup>8</sup>Institute of Theoretical Physics, University of Zurich, CH-8057 Zurich, Switzerland

Accepted 2014 January 17. Received 2014 January 16; in original form 2011 December 15

## ABSTRACT

Weak lensing by large-scale structure is an invaluable cosmological tool given that most of the energy density of the concordance cosmology is invisible. Several large ground-based imaging surveys will attempt to measure this effect over the coming decade, but reliable control of the spurious lensing signal introduced by atmospheric turbulence and telescope optics remains a challenging problem. We address this challenge with a demonstration that point spread function (PSF) effects on measured galaxy shapes in the Sloan Digital Sky Survey (SDSS) can be corrected with existing analysis techniques. In this work, we co-add existing SDSS imaging on the equatorial stripe in order to build a data set with the statistical power to measure cosmic shear, while using a rounding kernel method to null out the effects of the anisotropic PSF. We build a galaxy catalogue from the combined imaging, characterize its photometric properties and show that the spurious shear remaining in this catalogue after the PSF correction is negligible compared to the expected cosmic shear signal. We identify a new source of systematic error in the shear–shear autocorrelations arising from selection biases related to masking. Finally, we discuss the circumstances in which this method is expected to be useful for upcoming ground-based surveys that have lensing as one of the science goals, and identify the systematic errors that can reduce its efficacy.

**Key words:** gravitational lensing: weak – techniques: image processing – surveys – cosmology: observations.

## 1 INTRODUCTION

Modern cosmologists can simulate the invisible implications of modern cosmological models (e.g. those that can explain the cosmic microwave background, including Komatsu et al. 2011) to what is generally agreed to be a high level of precision (and probably accuracy; cf. Lawrence et al. 2010). The easily observable consequences of these models for observations of galaxies are not so easy to calculate (e.g. Rudd, Zentner & Kravtsov 2008; Conroy & Wechsler 2009; Simha et al. 2012), involving as they do the physics of the familiar but nevertheless stubbornly complicated baryons. Most of the precisely calculable components of these models –

namely the properties of the distribution of dark matter on large scales in relatively linear structures – are not readily observable.

For the foreseeable future, the most direct observation of these dark components is the measurement of the gravitational effects of dark structures on the images of distant background galaxies. These measurements are made almost exclusively via statistical estimation of the distortions in the ellipticities of background galaxies. This takes advantage of the fact that galaxies have no preferred orientation in a homogeneous, isotropic universe.

Lensing measurements have played a significant role in observational astrophysics in the last two decades, over a range of scales and physical regimes. Studies of galaxy evolution benefit from the ability to understand the dark matter haloes that host galaxies (e.g. Hoekstra, Yee & Gladders 2004; Hoekstra et al. 2005; Heymans et al. 2006; Mandelbaum et al. 2006a,b, 2009; Leauthaud et al. 2012). Cosmologists have no other way to directly map the

★E-mail: [huff.791@osu.edu](mailto:huff.791@osu.edu)

large-scale matter distribution, which is crucial for constraining models of dark energy and modified gravity (Zhang et al. 2007; Reyes et al. 2010). On small scales, maps of the matter distribution can be tied directly to tests of the cold dark matter paradigm and simulations of the formation and evolution of dark matter haloes.

Much has been made of the scientific potential of this technique. Five years ago, weak lensing was identified by the Dark Energy Task Force (Albrecht et al. 2006) as the most promising tool for constraining cosmological models. Several large ground-based and space-based survey proposals place a weak lensing measurement among their primary science drivers, including the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS),<sup>1</sup> the Dark Energy Survey (DES),<sup>2</sup> the Hyper Suprime-Cam (HSC; Miyazaki et al. 2006) survey, the Large Synoptic Survey Telescope (LSST),<sup>3</sup> *Euclid*<sup>4</sup> and the Wide-Field Infrared Survey Telescope (WFIRST).<sup>5</sup>

For all the promise, the technical challenges for these future experiments remain formidable. An order-unity distortion to background galaxy images is produced by a physical, projected matter overdensity of

$$\Sigma_{\text{crit}} = \frac{c^2}{4\pi G} \frac{d_s}{d_L d_{LS}}, \quad (1)$$

where  $d_L$ ,  $d_s$  and  $d_{LS}$  are the angular-diameter distance from the observer to the lens and source, and from the lens to the source, respectively. For characteristic distances of approximately 1 Gpc, the critical surface density is  $0.1 \text{ g cm}^{-2}$ . Typical fluctuations in the matter density field projected over cosmological distances are a thousand times smaller than this, so order 10-Mpc-scale density fluctuations in the universe will typically produce changes in galaxy ellipticities of the order of  $e \approx 10^{-3}$ – $10^{-2}$  in magnitude. In the shot-noise-dominated regime, the leading-order contribution to the variance in the correlation function of the ellipticity distortions is

$$\text{Var}(\xi_e) = \frac{\sigma_e^4}{N_{\text{pair}}^2}. \quad (2)$$

For a shallow ( $\langle z \rangle = 0.5$ ) galaxy survey with shape noise due to random galaxy ellipticities  $\sigma_e \approx 0.3$  and  $100 \text{ deg}^2$  of sky coverage, reducing the shot-noise contribution below the expected cosmological signal requires a surface density of usable source galaxies of at least a few per square arcminute.

Worse, for ground-based imaging surveys, the observed shape distortions arising from atmospheric turbulence and optical distortions from the telescope are typically of the order of several per cent, with coherence over angular scales comparable to that of the lensing shape distortions. A competitive measurement of the amplitude of matter fluctuations requires suppressing or modelling these coherent spurious distortions to the order of one part in  $10^3$ , and future surveys will need to do a factor of several better.

Achieving both the statistical precision and control of systematic errors that is required for such a measurement has proved to be a challenge. The early detections (Bacon, Refregier & Ellis 2000; Van Waerbeke et al. 2000; Rhodes, Refregier & Groth 2001; Hoekstra et al. 2002; Brown et al. 2003; Jarvis et al. 2003) showed the promise of the method and confirmed the existence

of lensing by a large-scale structure at roughly the expected level. However, they also highlighted some of the systematic errors: in particular, *B*-mode shear (which cannot be produced by lensing at linear order and is thus indicative of systematic effects) was present at a sub-dominant but non-negligible level. Since then, the weak lensing community has moved in the direction of both deep/narrow surveys with the *Hubble Space Telescope* (*HST*) and wide/shallow surveys on the ground. The Cosmological Evolution Survey (COSMOS) is the premier example of the former: in addition to two-point statistics (Massey et al. 2007a; Schrabback et al. 2010), it has also produced three-dimensional maps of the matter distribution (Massey et al. 2007a) and the lensing three-point correlation function (Semboloni et al. 2011). Excellent control of lensing systematics in COSMOS was also achieved thanks to the small number of degrees of freedom controlling the point spread function (PSF; mostly focus variation; Rhodes et al. 2007) and detailed modelling of charge transfer inefficiency (Massey et al. 2010). However, COSMOS covers only  $1.6 \text{ deg}^2$ , and the small field of view of *HST* instruments makes significantly larger surveys impractical. The principal recent ground-based cosmic shear programme has been the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS). There are now several cosmic shear results from different subsets of the CFHTLS data (Hoekstra et al. 2006; Semboloni et al. 2006; Benjamin et al. 2007; Fu et al. 2008), and the CFHT lensing team is completing a reanalysis using recent advances in PSF determination and galaxy shape measurement.

In light of the efforts shortly to be made by large, expensive surveys to measure cosmic shear, we consider it imperative to show that such a measurement can be performed accurately, without significant contaminating systematic errors, from a ground-based observatory. This goal includes doing a cosmic shear measurement with each of the wide-angle optical surveys that presently exist. To this end, we have re-coadded the repeat observations on the equatorial stripe (Stripe 82) of the Sloan Digital Sky Survey (SDSS), using methodology that will optimize these new co-adds for precision shear measurements. Our goal is to reduce the systematic errors arising from uncorrected PSF anisotropies below the statistical errors. We begin by specifying our requirements in Section 2, and describing the data that we use in Section 3. A description of the co-addition and catalogue-making pipeline follows in Section 4. We describe our method for estimating two-point functions of star and galaxy shapes in Section 5. Demonstrations of the data quality and suitability for sensitive weak lensing measurements are described in Section 6. We conclude with lessons for future experiments in Section 7.

## 2 DESIGN REQUIREMENTS

Weak lensing measurements on large scales are vulnerable to a variety of systematic measurement errors. In order to measure cosmic shear on the scales described above, we must first have a clear idea of what the possible sources of these systematic errors are, and to what level (quantitatively) they must be suppressed. This section describes in turn the most common generic sources of measurement error relevant for weak lensing, and lays out quantitative methods for detecting their presence in our final catalogue.

The PSF<sup>6</sup> of the SDSS exhibits significant spatial and temporal variations across the entire survey. We model these effects as a

<sup>1</sup> <http://pan-starrs.ifa.hawaii.edu/public/>

<sup>2</sup> <http://www.darkenergysurvey.org/>

<sup>3</sup> <http://www.lsst.org/>

<sup>4</sup> <http://sci.esa.int/euclid/>

<sup>5</sup> <http://wfirst.gsfc.nasa.gov/>

<sup>6</sup> Here we use the term ‘PSF’ to denote all contributions: the atmosphere, optics, tracking errors, charge diffusion and pixelization.

spatially varying convolution kernel  $G$ . The observed image  $I(\mathbf{x})$  at some position  $\mathbf{x}$  is related to the ‘true’ image  $f$  by

$$I(\mathbf{x}) = \int f(\mathbf{y})G(\mathbf{x} - \mathbf{y})d^2\mathbf{y}, \quad (3)$$

where  $G$  is the convolution kernel appropriate to the region of sky under examination.

One effect of a spatially varying PSF  $G$  is to produce a spurious shear field determined by the atmosphere and telescope that is statistically independent of and superposed upon the undistorted galaxy shape pattern. Point sources (stars and completely unresolved galaxies – we have no need, at present, to distinguish these) sample only the field sourced by  $G$ , and so can be used to constrain a model for the systematics field. Any uncorrected additive shear contribution due to the ellipticity of  $G$  will produce a correlation between the measured galaxy and point source shapes. This additive shear will be statistically uncorrelated with the true cosmic shear signal.

The masking steps of the catalogue-construction procedure can also produce a significant shape selection bias. For the photometric pipeline used here, masked regions are defined as sets of pixels; a galaxy is rejected from the catalogue if the set of pixels making up a galaxy intersects the set of masked pixels. On the masked region boundary, galaxies aligned across the mask boundary are more likely to be rejected from the catalogue than galaxies aligned along it producing a spurious shear. This will affect both stars and galaxies, but the effect on spurious galaxy ellipticities will be much larger than that on stars (as the dispersion in measured stellar ellipticities is very small). This mask selection bias produces an additive shear, which will also be statistically uncorrelated with the true cosmic shear signal.

These two effects enter together as an additive term in the shape clustering statistics, as

$$\xi_{\text{measured}}(\theta) = \xi_{\text{cosmic}}(\theta) + \xi_{\text{systematics}}(\theta). \quad (4)$$

The point source and galaxy populations have different sensitivities to the ellipticity of the PSF, to optical distortions and to the geometry of the mask. If these are accounted for then, as described in e.g. Bacon et al. (2003), a measurement of the point source–galaxy cross-correlation provides a straightforward estimate of the spurious signal sourced by uncorrected PSF variation.<sup>7</sup> We will require that the amplitude of this spurious correlation in our final shape catalogue be sub-dominant to the statistical errors – in particular, that the additive PSF systematics amplitude be constrained to less than the statistical errors.

The average ellipticity measured for the gravitationally sheared images of a population of galaxies is proportional to the applied shear; the exact value of this calibration depends on the surface brightness profiles of the galaxies. We will address the shear calibration uncertainties in a companion paper.

### 3 DATA

#### 3.1 The Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS; York et al. 2000) and its successor SDSS-II (Frieman et al. 2008) mapped 10 000 square degrees across the north galactic cap using a dedicated wide-field

2.5 m telescope at Apache Point Observatory in Sunspot, New Mexico (Gunn et al. 2006). The SDSS camera, described in Gunn et al. (1998), images the sky in five optical bands ( $u, g, r, i, z$ ; Fukugita et al. 1996; Smith et al. 2002) with the charge-coupled device (CCD) detectors reading out at the sidereal rate. Each patch of sky passes in sequence through the five filters (in the order  $r, i, u, g, z$ ) along one of the six columns of mosaicked CCDs, and is exposed once in each filter for 54.1 s. The site is monitored for photometricity (Hogg et al. 2001; Tucker et al. 2006). Data undergo quality assessment (Ivezić et al. 2004), and final calibration is done using the ‘ubercalibration’ procedure based on photometry of stars in run overlap regions (Padmanabhan et al. 2008). We use the data from the seventh SDSS data release (Abazajian et al. 2009), with an updated calibration from the subsequent data release.

The footprint of one night’s observing is six columns of imaging the width of one CCD (13.52 arcmin) separated by slightly less than one CCD width (11.65 arcmin). Imaging taken during a continuous period of time on one night is collectively termed a *run*; each separate column of imaging is, sensibly, a *camera column* (or ‘camcol’), and the imaging along each camera column is chopped for processing purposes into 8.98 arcmin long *frames* or *fields*. Successive runs are interleaved, in order to fill in the gaps between camera columns. Pairs of interleaved runs along the same great circle are *stripes* (each of which has a north and a south *strip*).

#### 3.2 Stripe 82

Most of the SDSS imaging data were acquired in the northern galactic cap, with galactic latitude  $|b| > 30$ . For commissioning, and during sidereal times when the primary survey region was unavailable, the telescope frequently imaged a  $2^\circ.5$  wide stripe of sky along the celestial equator with right ascension (RA) in the interval  $-50^\circ < \text{RA} < +50^\circ$ . The SDSS-II supernova project (Frieman et al. 2008) observed this region many times during the months of September–November over the years 2005–2007 to collect multicolour light curves of Type Ia supernovae. In the survey nomenclature, this region is Stripe 82. At any given location on the Stripe, there are on average 120 contributing interleaved imaging runs, comprising in aggregate almost as much imaging data as exist in the remainder of the combined SDSS-I and SDSS-II footprint. It is here that significant gains can be made from image co-addition.

#### 3.3 Single-epoch data processing

The raw imaging data are processed by the automated SDSS photometric pipeline, *PHOTO* (Lupton et al. 2001). This pipeline has components to handle astrometric and photometric calibration as well as catalogue construction; it also generates an array of data quality measurements describing the telescope PSF, the locations of unreliable pixels and measurements of the photometric quality of individual frames. Many of these data quality indicators are used during the construction of the co-add imaging and its associated catalogue. Their use is described below. A detailed description of the image processing pipeline and its outputs can be found in Stoughton et al. (2002). Outputs can be found in locations specified by the SDSS data release papers (Abazajian et al. 2003, 2004, 2005, 2009; Adelman-McCarthy et al. 2006, 2007, 2008).

*PHOTO* produces a number of intermediate outputs for the single-epoch SDSS imaging that we use in the co-addition process. Corrected frames, or *fpC* files, are produced by the pipeline from the raw CCD images of single frames; these are bias-subtracted and flat-fielded, and a non-linearity correction is applied where appropriate.

<sup>7</sup> This statement is true for sufficiently large areas that any chance superpositions of PSF ellipticity and the lensing shear average out. For this reason, we impose this test on chunks with area  $\geq 25 \text{ deg}^2$ .

These are the images that are combined during the co-addition process below.

PHOTO also generates a bitmask (an `fpm` file) for each frame describing pixels that are known to be defective. Pixels are marked in this bitmask as saturated, cosmic ray contaminated, interpolated (if a column or pixel is known to be saturated, or is a priori marked as unreliable, PHOTO interpolates over that region). We use these bitmasks to exclude bad pixels from the image co-addition.

Astrometric solutions (asTrans files) are produced by ASTROM for each SDSS frame. Systematic errors in the astrometric positioning are reported to less than 50 mas, and the relative astrometry between successive overlapping frames is approximately 10 mas (Pier et al. 2003). The astrometric solution for each run (Pier et al. 2003) is determined by matching against astrometric standard stars from the United States Naval Observatory (USNO) CCD Astrograph Catalogue (Zacharias et al. 2000). The co-addition algorithm relies on the astrometric solutions provided; we have found it unnecessary to resolve the astrometry.

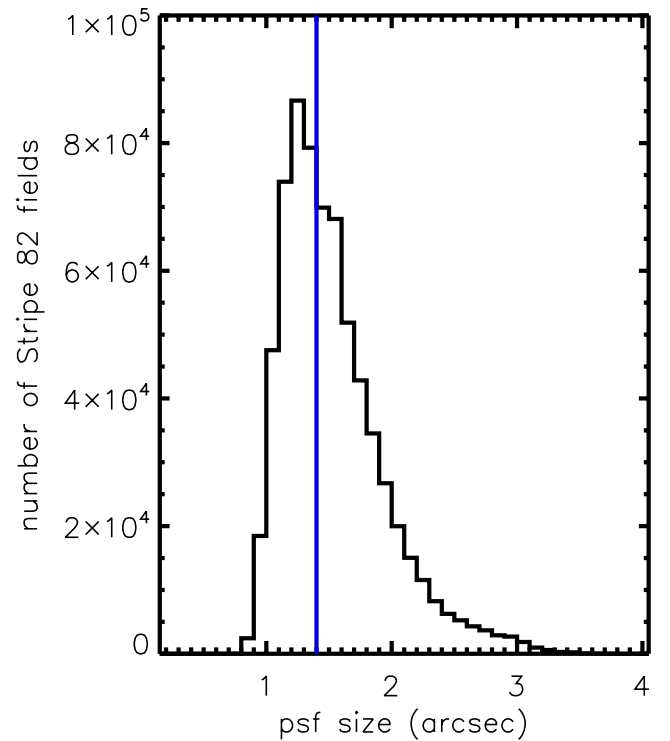
For photometric calibration, we use the ‘ubercal’ solutions derived by Padmanabhan et al. (2008).

The SDSS pipeline uses bright, isolated stars with apparent magnitudes brighter than 19.5 to construct a model of the PSF and its variation across each frame. For each frame, the stellar images for the three neighbouring frames along the scan in both directions are used to produce a set of Karhunen–Loève (KL) eigenimages describing the PSF variation (Lupton et al. 2001). A global PSF model for the frame is constructed by allowing the first few KL components to vary up to second order in the image coordinates across the frame, with the coefficients of the variation being tied to the aforementioned bright stars. The KL eigenimages and coefficients of their spatial variation are stored by PHOTO for each band in the `psField` files. These are taken as inputs to the co-addition process and used for PSF correction. We will test the fidelity of this PSF model in the co-added images on stars that were too faint to perform a reliable PSF determination in the single-epoch data.

## 4 ALGORITHMS

Our general strategy for correcting for the effects of seeing is similar to that suggested in Bernstein & Jarvis (2002). We will apply a rounding kernel to each single-epoch image prior to stacking the ensemble. In the presence of a perfectly unbiased shape-measurement method, the application of the rounding kernel will unnecessarily destroy information. At the present time, however, no unbiased shape-measurement method is known to exist (Kitching et al. 2012). Assessments of the performance of shear measurement algorithms on simulated images (e.g. Massey et al. 2007b; Melchior et al. 2010) have found that measurement biases frequently depend on the form of the PSF. As long as the effect of the PSF on unsmear galaxy images can be described as a convolution, the rounding kernel can be designed to adjust the image PSF to a form that is convenient for the shape-measurement method at hand. This comes at a cost in potential statistical power, however, as the size of the PSF can only be increased.

The large variation in SDSS PSF sizes (see Fig. 1) will require a trade-off between rejection of a large fraction of the available imaging and significant dilution of the signal due to the rounding convolution. Stacking the images without a kernel, however, will produce a PSF with large variations – including *steps* at run boundaries or the edges of regions masked due to e.g. cosmic rays – that will be difficult to model accurately.



**Figure 1.** The distribution of PSF full width at half-maximum (FWHM) in the *r* band for all frames on Stripe 82. The half-width of the target PSF after rounding is indicated by the solid vertical line.

### 4.1 Field smoothing

This section describes the operation of smoothing the map so as to make the effective PSF equal to some target PSF. Here we will denote the intrinsic PSF of the telescope by  $G(\mathbf{x})$ , so that if the intrinsic intensity of an object on the sky is  $f(\mathbf{x})$ , the actual image observed is

$$I(\mathbf{x}) = \int G(\mathbf{y})f(\mathbf{x} - \mathbf{y})d^2\mathbf{y} \equiv [G \otimes f](\mathbf{x}). \quad (5)$$

Of course, this image is only sampled at values of  $\mathbf{x}$  corresponding to pixel centres. Our principal objective here is to construct the kernel  $K$  such that

$$[I \otimes K](\mathbf{x}) = [\Gamma \otimes f](\mathbf{x}) \text{ or } [G \otimes K](\mathbf{x}) = \Gamma(\mathbf{x}), \quad (6)$$

where  $\Gamma$  is the target PSF. In order to do this, we need to first choose a target PSF  $\Gamma$  and then determine the appropriate convolution kernel  $K$ , which will differ for every imaging run contributing to the co-adds at a given position depending on the full position-dependent PSF model in each run. These are the subjects of Sections 4.1.1 and 4.1.2, respectively.

#### 4.1.1 The target PSF

Here we consider the target PSF  $\Gamma$ . It must be constant across different runs in order for the co-add procedure to make sense, although it need not be the same in different filters. There is a large advantage in having  $\Gamma$  be circularly symmetric. Gaussians are convenient since most galaxy shape-measurement codes are based on Gaussian moments, but this is not a requirement. In fact the PSF  $G$  delivered by most telescopes, including the SDSS, has ‘tails’ due to the atmosphere at large radius that are far above what one could be expected from a Gaussian. These can in principle be removed



by a convolution kernel  $K$  that has negative tails at large radius, but there are problems when these tails extend to the field boundaries or across bad columns in the CCD. Therefore, we have chosen the double-Gaussian form for  $\Gamma$ :

$$\Gamma(\mathbf{x}) = \frac{1 - f_w}{2\pi\sigma_1^2} e^{-x^2/2\sigma_1^2} + \frac{f_w}{2\pi\sigma_2^2} e^{-x^2/2\sigma_2^2} \quad (7)$$

with  $\sigma_2 > \sigma_1$ . This functional form manifestly integrates to unity, and has a fraction  $f_w$  of the light in the ‘large’ Gaussian. The two Gaussians have widths  $\sigma_1$  and  $\sigma_2$ , respectively, with  $\sigma_1 < \sigma_2$ .

The parameters of the double-Gaussian were adjusted by trial and error so that a compactly supported kernel  $K$  ( $13 \times 13$  pixels) can achieve  $G \otimes K \approx \Gamma$  for a wide range of real PSFs  $G$  delivered by the SDSS. The most critical parameter is the width of the central Gaussian,  $\sigma_1$ . This is the main parameter controlling the seeing of the final co-added image: if it is set too high, then many galaxies become unresolved, whereas if it is set too low, then a large number of fields with moderate seeing will have to be rejected because it will be impossible to find a kernel  $K$  that achieves the target PSF without dramatically amplifying the noise.

The PSF size distribution in the  $r$  band is shown in Fig. 1.

#### 4.1.2 The convolution kernel and its application

Equation (6) can formally be solved in Fourier space by taking the ratio,  $\tilde{K}(\mathbf{k}) = \tilde{\Gamma}(\mathbf{k})/\tilde{G}(\mathbf{k})$ , where the tilde denotes the Fourier transform and  $\mathbf{k}$  the wave vector. Unfortunately, this simple idea comes with two well-known problems. One is that if the PSF has power only up to a certain wavenumber  $k_{\max}$ , then it is impossible to divide by  $\tilde{G}(\mathbf{k}) = 0$ . The other is that the PSF varies slowly across the field, i.e.  $G$  in equation (6) formally depends on  $\mathbf{x}$  as well as  $\mathbf{y}$ .

The solution to the first problem is that instead of taking a simple ratio in Fourier space, we minimize the  $L^2$  norm of the error,

$$\mathcal{E}_1 = \int |\Gamma(\mathbf{x}) - [G \otimes K](\mathbf{x})|^2 d^2\mathbf{x} \equiv \|\Gamma - G \otimes K\|^2, \quad (8)$$

subject to a constraint on the  $L^2$  norm of the kernel:

$$\mathcal{E}_2 = \int |K(\mathbf{x})|^2 d^2\mathbf{x} \equiv \|K\|^2. \quad (9)$$

If the input noise is white (which is a good approximation), then the noise variance on an individual pixel in the convolved image is  $\mathcal{E}_2$  times the noise variance in the input image. Roughly speaking, for kernels that attempt to ‘deconvolve’ the original PSF, and consequently have large positive and negative contributions,  $\mathcal{E}_2$  will come out to be very large. We adopt a requirement that  $\mathcal{E}_2 \leq 1$ . For kernels that poorly approximate the target PSF  $\Gamma$ ,  $\mathcal{E}_1$  will be very large. The problem of minimizing  $\mathcal{E}_1$  subject to a constraint on  $\mathcal{E}_2$  is most easily solved by transforming to the Fourier domain and then using the method of Lagrange multipliers:

$$\tilde{K}(\mathbf{k}) = \frac{\tilde{G}^*(\mathbf{k})\tilde{\Gamma}(\mathbf{k})}{|\tilde{G}(\mathbf{k})|^2 + \Lambda}. \quad (10)$$

Here the positive real number  $\Lambda$  is the Lagrange multiplier and its value is adjusted until  $\mathcal{E}_2 = 1$ .  $\Lambda$  plays the role of regulating the deconvolution; indeed one can see that for Fourier modes present in the image,  $\tilde{G}(\mathbf{k}) \neq 0$ , we have  $\lim_{\Lambda \rightarrow 0^+} \tilde{K}(\mathbf{k}) = \tilde{\Gamma}(\mathbf{k})/\tilde{G}(\mathbf{k})$ .

To summarize, equation (10) finds the convolving kernel  $K$  that makes the final PSF  $G \otimes K$  as close as possible (in the least-squares sense) to  $\Gamma$  without amplifying the noise. The kernel is truncated into a  $13 \times 13$  pixel region centred at the origin in order to avoid boundary effects and to prevent problems such as bad

columns, saturated stars or cosmic rays from ‘leaking’ all over the field. We also re-scale the resulting kernel to integrate to unity ( $\tilde{K}(\mathbf{0}) = 1$ ) but since  $\Lambda$  is small, typically of the order of  $10^{-5}$ , this has no practical effect. Note that since  $G(\mathbf{x})$  and  $\Gamma(\mathbf{x})$  are both real functions, it follows that in Fourier space they satisfy the conditions  $\tilde{G}(\mathbf{k}) = \tilde{G}^*(-\mathbf{k})$  and  $\tilde{\Gamma}(\mathbf{k}) = \tilde{\Gamma}^*(-\mathbf{k})$ , and then equation (10) guarantees that a similar condition holds for  $K$ : the convolution kernel  $K(\mathbf{x})$  is real.

The second problem – the variation of the PSF across the field – is handled by taking the reconstructed PSF on a grid of  $8 \times 6$  points separated by 298 pixels (2 arcmin) in each direction, and constructing a grid of 48 kernels  $K$ . The kernels are then interpolated bilinearly between the four nearest grid points, and then the final image  $F(\mathbf{x})$  is constructed according to

$$F(\mathbf{x}) = \int K_x(\mathbf{y}) I(\mathbf{x} - \mathbf{y}) d^2\mathbf{x}, \quad (11)$$

where  $K_x$  is the kernel reconstructed at position  $\mathbf{x}$  in the field.

The convolution kernel will not capture PSF model fluctuations on scales below 2 arcmin. Since the SDSS model PSFs are quadratic functions over the chip, features at the arcminute scale and smaller are not captured anyway. We show below that, even at  $\theta = 1$  arcmin, the remaining PSF variations not captured by the kernel are very small compared to the expected shot-noise errors in the two-point statistics at those scales.

Obviously there will be instances in which the kernel reconstruction is not good enough. Therefore, a set of cuts must be applied to the resulting kernels.<sup>8</sup> In order to construct these cuts, we consider the Gaussian-weighted moments of the residual  $\Gamma - G \otimes K$ , i.e.

$$M_{\alpha\beta} = \frac{1}{\pi\sigma_1^2} \int [\Gamma - G \otimes K](\mathbf{x}) \frac{x_1^\alpha x_2^\beta}{\sigma_1^{\alpha+\beta}} e^{-x^2/2\sigma_1^2} d^2\mathbf{x}. \quad (12)$$

We require that the PSF-induced ellipticity be smaller than  $10^{-3}$  in order to ensure a clean cosmic shear signal with this characteristic strength. The cuts are as follows.

(1) We reject an entire field if the SDSS software used to determine the PSF (the postage stamp pipeline, or PSP) failed to determine a good PSF model in the single-epoch imaging, or was forced to use a low-order fit to the PSF (PSP\_STATUS != 0).

(2) We reject the cases where the PSF residual is too large regardless of the moments, i.e.

$$\frac{\|\Gamma - G \otimes K\|^2}{\|\Gamma\|^2} > \text{CUT\_L2}. \quad (13)$$

This cut is intended to guard against pathological cases, where the PSF that results from the rounding kernel passes the other cuts below, but is still a poor fit to the target. The value of this cut was determined by trial and error; on trial co-adds consisting of a small fraction of the Stripe 82 footprint, we applied the rounding kernel and measured the ellipticities of the resultant stars. The value of this cut was adjusted until the residual stellar ellipticities were reliably less than  $10^{-3}$ .

(3) We reject the cases where the Gaussian-weighted offset is more than CUT\_OFFSET  $\sigma_1$ , i.e.

$$\sqrt{M_{01}^2 + M_{10}^2} > \text{CUT\_OFFSET}. \quad (14)$$

This cut removes the cases where significant astrometric offsets were introduced during stacking. The ellipticity errors induced

<sup>8</sup> It should be noted that residual anisotropies from the Lagrange multiplier  $\Lambda$  are affected by the quality cuts described below.

**Table 1.** Parameters for the PSF repair in different filters.

Parameter	<i>u</i>	<i>g</i>	<i>r</i>	<i>i</i>	<i>z</i>	Units
Target PSF parameters						
$\sigma_1$ (PSF_SIZE)	1.80	1.40	1.40	1.40	1.40	pixels
$\sigma_2$ (PSF_SIZE_WING)	5.10	5.10	5.10	5.10	5.10	pixels
$f_w$ (FRACWING)	0.035	0.035	0.035	0.035	0.035	pixels
FWHM of target PSF $\Gamma$	1.68	1.31	1.31	1.31	1.31	arcsec
50 per cent encircled energy radius	0.86	0.67	0.67	0.67	0.67	arcsec
Kernel acceptance parameters						
CUT_L2	0.001	0.0025	0.0025	0.0025	0.0025	
CUT_OFFSET	0.04	0.01	0.01	0.01	0.01	
CUT_ELLIP	0.002	0.0005	0.0005	0.0005	0.0005	
CUT_SIZE	0.01	0.0025	0.0025	0.0025	0.0025	
CUT_PROF4	0.04	0.01	0.01	0.01	0.01	
Co-addition parameters						
DELTA_SKY_MAX1	0.5	0.25	0.25	0.25	0.25	nmgy arcsec <sup>-2</sup>
DELTA_SKY_MAX2	0.04	0.02	0.02	0.02	0.02	nmgy arcsec <sup>-2</sup>

by such offsets scale as the square of the astrometric offset, with an order-unity proportionality constant (Bernstein & Jarvis 2002). With this assumption, the choice of cut here reduces the centroiding error in the ellipticity below our fiducial ceiling of  $10^{-3}$ .

(4) We reject the cases where the ellipticity of the final PSF exceeds CUT\_ELLIP, i.e.

$$\sqrt{(M_{02} - M_{20})^2 + (2M_{11})^2} > \text{CUT\_ELLIP}. \quad (15)$$

The quantity CUT\_ELLIP is a non-adaptive ellipticity measured with a circular Gaussian weight  $\sigma_1$ . In the absence of the small-amplitude component of our PSF (i.e.  $f_w = 0$ ), and for small residuals, this quantity would be equal to half the adaptive moment ellipticity; our chosen cut corresponds to our fiducial ellipticity ceiling of  $10^{-3}$ .

(5) We reject the cases where the PSF size error exceeds CUT\_SIZE, i.e.

$$|M_{02} + M_{20} - M_{00}| > \text{CUT\_SIZE}. \quad (16)$$

This cut limits the effect of PSF dilution correction errors (i.e. errors in  $R_2$ ; see Section 4.9 below). The effect of the dilution correction on the measured ellipticity is smaller than that of the PSF ellipticity by a factor of the galaxy resolution. This cut ensures that dilution correction errors on the ellipticity are below  $10^{-3}$  for galaxies near our adopted resolution limit  $R_2 = 0.333$ .

(6) We reject the cases where the radial profile of the PSF is severely in error as determined by the fourth moment, i.e.

$$|M_{40} + 2M_{22} + M_{04} - 2M_{00}| > \text{CUT\_PROF4}. \quad (17)$$

The value of this cut was also determined by trial and error, as part of the same procedure as with CUT\_L2 above.

The specific values of the parameters chosen for the cuts depend on the band and are shown in Table 1. The tightest constraints on the quality of the PSF are in *g*, *r*, *i* and *z* bands (*r* and *i* are used to measure galaxy shapes). In the *u* band, where the average image quality is much lower than in the other bands, more liberal cuts can be applied because we are interested primarily in the total flux, not the shape. Also there is more to gain from liberal cuts because the signal-to-noise ratio in the *u* band is lower. Nevertheless, a serious error in the size of the PSF will result in erroneous photometry, and spurious ellipticity could introduce colour/photo-*z* or selection biases that depend on galaxy orientation, so some cuts must be applied.

## 4.2 Noise symmetrization

It is a well-known fact in weak lensing that even if the PSF in an image has been corrected to have perfectly circular concentric isophotes, it is possible to produce spurious ellipticity if there is anisotropic correlated noise. For example, if the PSF is elongated in the  $x_1$  direction and is ‘fixed’ by smoothing in the  $x_2$  direction, the resulting map has more correlations in the  $x_2$  direction than in  $x_1$ . This can lead to (1) centroiding biases, in which the error on the galaxy centroid is larger in the  $x_2$  than the  $x_1$  direction, thus yielding more galaxies that appear aligned in the  $x_2$  than  $x_1$  direction; and (2) biases in which noise fluctuations tend to be elongated in the  $x_2$  direction, so that positive noise fluctuations on top of a galaxy (which increase its likelihood of detection) tend to make it aligned in the  $x_2$  direction whereas negative fluctuations (which decrease the likelihood of detection) make the galaxy aligned in the  $x_1$  direction. For a detailed description of noise-induced ellipticity biases, see Kaiser (2000) or Bernstein & Jarvis (2002). These phenomena can all mimic lensing signals and hence should be eliminated from the data. Our method of doing this is to add synthetic noise to each field so as to give the noise properties 4-fold rotational symmetry. To be precise, we want the power spectrum of the total noise (actual plus synthetic) to satisfy

$$P_N(\mathbf{k}) = P_N(\mathbf{e}_3 \times \mathbf{k}), \quad (18)$$

where  $\mathbf{e}_3$  is a vector normal to the plane of the image; the cross product operation  $\mathbf{e}_3 \times$  rotates a vector by  $90^\circ$ . Even though it is not circularly symmetric, this is sufficient to guarantee zero mean ellipticity for a population of randomly oriented galaxies because ellipticity reverses sign under  $90^\circ$  rotations.<sup>9</sup> In principle, *m*-fold symmetry for any integer  $m \geq 3$  would suffice; however, 4-fold symmetry is the only practical possibility for a camera with square pixels. For obvious reasons, we would like to achieve this by adding the minimal amount of synthetic noise possible.

<sup>9</sup> In group theory language, the noise properties are symmetric under the 4-fold rotation group  $C_4$ , which is a subgroup of the full rotations  $SO(2)$ . The condition for zero mean ellipticity due to noise is that ellipticity fall into one of the non-trivial representations of the noise symmetry group.

The simplest way to achieve equation (18) is to decompose the power spectrum into its actual ('act') and synthetic ('syn') components:

$$P_N(\mathbf{k}) = P_N^{(\text{act})}(\mathbf{k}) + P_N^{(\text{syn})}(\mathbf{k}). \quad (19)$$

The actual component is the white noise variance  $v$  in the input image, smoothed by the convolving kernel:

$$P_N^{(\text{act})}(\mathbf{k}) = v |\tilde{K}(\mathbf{k})|^2. \quad (20)$$

Since  $K$  is real, this power spectrum has 2-fold rotational symmetry:  $P_N^{(\text{act})}(\mathbf{k}) = P_N^{(\text{act})}(-\mathbf{k})$ . The minimal synthetic noise power spectrum that satisfies equation (18) is then

$$P_N^{(\text{syn})}(\mathbf{k}) = \max \left[ P_N^{(\text{act})}(\mathbf{e}_3 \times \mathbf{k}) - P_N^{(\text{act})}(\mathbf{k}), 0 \right]. \quad (21)$$

Gaussian noise with this spectrum can be obtained by taking its square root,

$$\tilde{T}(\mathbf{k}) = \sqrt{P_N^{(\text{syn})}(\mathbf{k})}, \quad (22)$$

and transforming to configuration space  $T(\mathbf{x})$ . Then one generates white noise with unit variance and convolves it with  $T$ . Since the PSF and hence  $K$  varies across the field,  $T$  must also vary; its value is interpolated from the same  $8 \times 6$  grid of reference points as used for  $K$ .

The Gaussian white noise was generated using Numerical Recipes `gasdev` modified to use the `ran2` uniform deviate generator (Press et al. 1992). The seed was chosen by a formula based on the run, `camcol`, field number and filter to guarantee that the same seed was never used twice in the reductions, and that the same sequence will be generated if the software is re-run. For each field, a sequence of  $2048 \times 1361$  Gaussian deviates is generated; since there are 128 rows of overlap between successive fields, we fill in the last 128 rows of each field with the first  $2048 \times 128$  deviates from the next field. It is also essential that the period of the generator be longer than the total number of pixels in the survey (of the order of a few  $\times 10^{12}$ ), a requirement which is *not* fulfilled by many generators, since otherwise the same synthetic 'noise' pattern will repeat itself throughout the survey.

The image  $F(\mathbf{x})$  after addition of the synthetic noise is a `kImage`.

### 4.3 Single-image masking

Once the kernel-convolved, noise-added image (`kImage`) is constructed for each run that will contribute to the co-adds at a given position, it is necessary to construct a mask before co-addition. The mask must remove the usual image defects as well as diffraction spikes. It is constructed as described in this section, and is termed the `kMask`.

We begin by masking out all pixels in  $F(\mathbf{x})$  for which the convolution (equation 11) integrates over a bad pixel. Since  $K$  has compact support – it is non-zero only in a  $13 \times 13$  pixel region – this means that for each bad pixel in  $I(\mathbf{x})$  we mask out a  $13 \times 13$  block in  $F(\mathbf{x})$ . Our definition of a 'bad pixel' is one that is out of the field; was interpolated by `PHOTO` (usually due to being in a bad column); is saturated; is potentially affected by ghosting (via the `fpM` ghost flag); was not checked for objects by `PHOTO`; is determined by `PHOTO` to be affected by a cosmic ray; or had a model subtracted from it. Note that the first cut means that a six-pixel region is rejected around the edge of the field.

The second and more sophisticated mask is applied to remove diffraction spikes from stars. The secondary support structure responsible for the diffraction spikes is on an altitude–azimuth mount,

so that the diffraction spikes appear at position angles of  $45^\circ$ ,  $135^\circ$ ,  $225^\circ$  and  $315^\circ$  in the altitude–azimuth coordinate system. Therefore, in the equatorial runs, the orientation of the diffraction spikes relative to equatorial coordinates changes depending on the hour angle of observation. If no correction for this is made, then after co-addition of many runs, even moderately bright stars have a hedgehog-like pattern of diffraction spikes at many position angles that can affect a significant fraction of the area.

Our procedure for removing diffraction spikes is as follows. We first identify objects with a PSF flux (i.e. flux defined by a fit to the PSF) exceeding some threshold (corresponding to  $9.7 \times 10^5$ ,  $8.5 \times 10^5$ ,  $2.2 \times 10^5$ ,  $1.7 \times 10^5$  and  $1.1 \times 10^6$  electrons in filters  $r$ ,  $i$ ,  $u$ ,  $z$  and  $g$ , respectively). Around these objects, we mask a circle of radius 20 pixels (8 arcsec) and four rectangles of width 8 pixels (3 arcsec) and length 60 pixels (24 arcsec). The rectangles have the object centroid at the centre of their short side, and their long axis extends radially from the centroid in the direction of the expected diffraction spike.

### 4.4 Resampling

In order to co-add images, we must first resample them into a common pixelization. Ideally, we would like this pixelization to be both conformal (no local shape distortion) and equal-area (convenient for total flux measurements). Unfortunately because the sky is curved, it is impossible to achieve both of these conditions. However, since our analysis uses a narrow range of declinations around the equator ( $|\delta| \leq 1.3$ ), we can come very close by choosing a cylindrical projection; the obvious choice is Mercator (perfectly conformal) or Lambert (perfectly equal-area). In our case, the Mercator projection would result in the pixel scale being different by  $\Delta\theta/\theta = 2.6 \times 10^{-4}$  at the Equator versus at  $\delta = \pm 1.3$ . (The area error is twice this, or  $5.2 \times 10^{-4}$ .) The Lambert projection would preserve shapes at the Equator but the coordinate system would have a shear of  $\gamma = 2.6 \times 10^{-4}$  at  $\delta = \pm 1.3$ . Neither of these problems is particularly serious, since either could if necessary be corrected in the flux or shape measurements. We have chosen the Mercator projection because the small cosmic shear signal means that we are much more sensitive to a given percentage error in shear than in flux. Also, a flux error of  $5.2 \times 10^{-4}$  is insignificant compared to the error in the flat-fields, so there is no point in eliminating it at the expense of complicating the shear analysis.

The scale of the resampled pixels must be smaller than the native pixel scale on the CCD ( $\sim 0.396$  arcsec) in order to preserve information. However, it is desirable for it not to be too small, since this increases the data volume with no increase in information content. It should also not be nearly equal to the CCD scale in order to avoid production of a moiré pattern with large-scale power. We have used 0.36 arcsec.

The actual resampling step requires us to interpolate the image from the native pixelization on to the target Mercator pixelization. This is done by 36-point second-order polynomial interpolation on the  $6 \times 6$  grid of native pixels surrounding the target pixel.<sup>10</sup> A target pixel is considered masked if any of the 36 surrounding pixels are masked.

<sup>10</sup> Polynomial interpolation on an equally spaced grid of points converges to sinc interpolation in the limit that the number of grid points is taken to infinity. This is easily seen from the Lagrange interpolation formula and the infinite product,  $\prod_{n=1}^{\infty} (1 - x/n)(1 + x/n) = \sin(\pi x)/(\pi x)$ .

#### 4.5 Addition of images

After resampling the images, the next step is to combine them to produce the co-add. The combination proceeds in three steps: comparison of images to reject ‘bad’ regions that were not masked in earlier stages of the analysis; relative sky estimation; and stacking. Note that bad regions must be explicitly rejected: ‘robust’ algorithms such as the median are non-linear and slightly biased, and result in a final co-added PSF that depends on object flux and morphology, which is not acceptable for lensing studies.

Rejection of bad regions is critical because it is possible for some serious defects such as satellite trails to ‘leak through’ earlier stages of the analysis and not be `kMasked`. Rejection at this stage is also the best way to eliminate Solar system objects, most of which will be known, but which are not easily identified in the single-epoch `fpCs` but of course will not show up at the same coordinates in successive runs. We first bin each input image into  $4 \times 4$  resampled pixels. We then compare the binned images and reject the brightest or faintest image if it differs by more than `DELTA_SKY_MAX1` from the mean. When this rejection is done, we actually mask a  $20 \times 20$  resampled pixel region around the affected area. (We found that without this padding region, satellite trails were often incompletely masked because they passed through the corners of some  $4 \times 4$  regions and did not sufficiently affect the mean flux.)

Next we compute the difference in sky level among all of the  $N$  images.<sup>11</sup> This difference must be determined and removed because otherwise a masked pixel in an image with below average sky will appear as a bright spot in the co-added image. We compute the relative sky level – an often neglected step in co-addition – as follows. For each pair  $(i, j)$  of co-added images, we compute the difference map  $F_i - F_j$  and take the median in  $125 \times 125$  resampled pixel blocks. This is taken as an estimate of the sky difference  $S_i - S_j$ . From these differences, we obtain the unweighted least-squares solution for the sky levels  $\{S_i\}$ , up to an additive offset (the absolute sky level cannot be determined by this procedure). The mean of these levels is denoted by  $\bar{S} = \sum_{i=1}^N S_i / N$ . We add to the  $i$ th image the quantity  $\bar{S} - S_i$  interpolated to a particular point  $\mathbf{x}$  by four-point interpolation from the nearest block centres. An entire block is masked out if  $|\bar{S} - S_i| > \text{DELTA\_SKY\_MAX2}$  and if it is an extremal value (either the highest or lowest sky value).

The stacking of the images works by the usual formula

$$F_{\text{tot}}(\mathbf{x}) = \frac{\sum_{i=1}^N w_i(\mathbf{x}) F_i(\mathbf{x})}{w_{\text{tot}}(\mathbf{x})}, \quad (23)$$

where  $w_{\text{tot}}(\mathbf{x}) = \sum_{i=1}^N w_i(\mathbf{x})$  and  $w_i$  are the weights. Because the noise is correlated, the optimal weights are scale dependent; we have chosen the optimal weights in the limit of small  $k$ , i.e. large scales. That is,  $w_i = v^{-1}$  where  $v$  is the noise variance in image  $i$ . For photometry of large objects,  $w_{\text{tot}}$  can be thought of as an inverse white noise variance, i.e. the mean square noise flux in a region of area  $\Omega$  is  $1/w_{\text{tot}}\Omega$ . However, for small objects (which are always our concern), this is not the case and the error bars must be computed from the measured noise properties of the co-add.

<sup>11</sup> While the `fpC` images generated from single-epoch data by `PHOTO` are sky subtracted, in practice this initial sky subtraction was not sufficiently smooth to avoid the appearance of large background brightness variation in the co-add images. This should not be surprising, as `PHOTO` has known sky-subtraction problems (Aihara et al. 2011).

An example of a co-added image, and comparison to a single-epoch image, is shown in Fig. 2.

#### 4.6 Additional masking

Before constructing the photometric catalogues, we zero all pixels contaminated by bright stars in the Tycho catalogues (Høg et al. 2000), replacing them with random noise of appropriate amplitude. Pixels masked in this manner have the ‘INTERP’ bit set in the input `fpM` files, so that the downstream analysis can exclude objects that incorporate pixels from a masked region. Pixels that are `kMasked` (according to one of the above criteria) also have ‘INTERP’ bits set. This final step results in a catalogue with a complex geometry, which will be demonstrated explicitly in Section 4.9.

#### 4.7 Photometric catalogues

Once each co-added image is constructed, we detect objects using the catalogue-construction portion of the SDSS photometric pipeline, `PHOTO-FRAMES`. The details of `FRAMES`’s catalogue construction and object measurement process are described more fully elsewhere (Stoughton et al. 2002; Lupton et al., in preparation). It is nevertheless useful to review the important parts of the `FRAMES` algorithms.

`PHOTO-FRAMES` requires as input a set of long integer images, and a considerable array of inputs describing the properties of the telescope and the observing conditions. Principal among these is a description of the telescope PSF. For single-epoch data, `FRAMES` uses a principal-component decomposition of the variation of the PSF across five adjacent fields. The components of this decomposition are allowed to vary as a polynomial (typically quadratic) in the image coordinates across each frame. As the co-added images have the same target PSF in every image, this target PSF is stored as the first principal component. For fast computation of object properties, the pipeline uses a double-Gaussian fit to the PSF; as this is the exact form of the target PSF resulting from the rounding kernel applied above, we simply use the target PSF parameters.

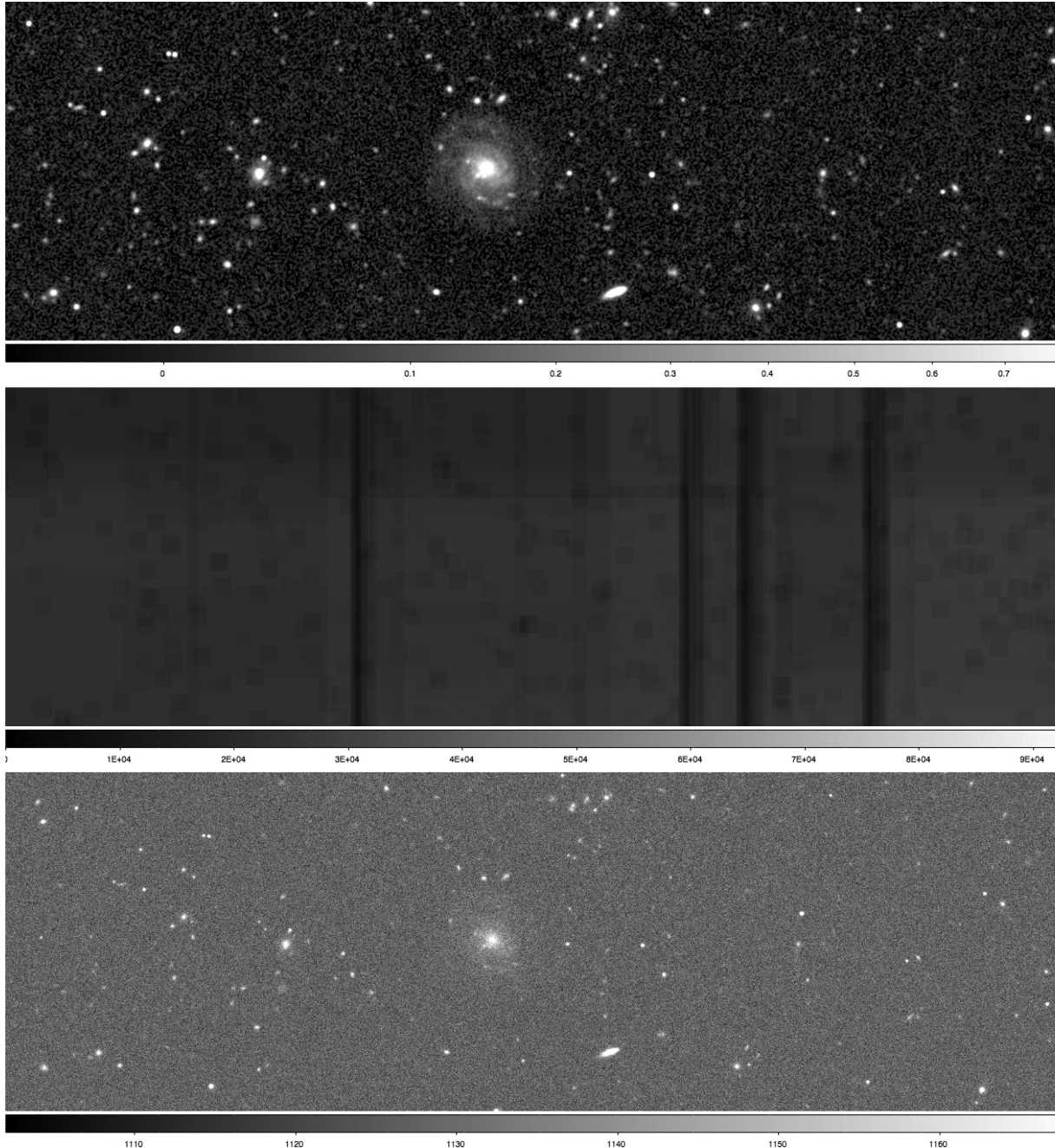
`FRAMES` first smooths the image with the narrower of the two Gaussian widths describing the PSF. Collections of connected pixels greater than seven times the standard deviation of the sky noise are marked as objects. Each object is grown by six pixels in each direction. For each object, the list of connected pixels is then culled of peaks less than three times the local standard deviation of the sky.

In order to avoid including objects that represent random noise fluctuations, catalogue galaxies are required to have statistically significant ( $>7\sigma$ ) detections in both the  $r$  and  $i$  bands. Note that this is a higher threshold than the  $>5\sigma$  cut used at this stage in the standard single-epoch SDSS processing. This was necessitated by the fact that the pixel noise in the `kImages` is correlated by the convolution process.

In the standard SDSS pipeline, `FRAMES` then re-bins the image and repeats the search. We choose not to use objects found in this manner, as the shape measurements of these very low surface brightness galaxies would not be reliable.

This detection algorithm is repeated in each filter separately. Objects detected in multiple bands are merged to contain the union of the pixels in each band if they overlap on the sky. The list of peak positions in each band is preserved. The centre of the resulting single object is determined by the location of the highest peak in the





**Figure 2.** A comparison of a co-added image (upper panel), its inverse variance map (middle panel) and a single-epoch input map (bottom panel). The co-added image is centred on RA  $01^{\text{h}}56^{\text{m}}34^{\text{s}}.8$ , Dec.  $-01^{\circ}10'35''$  (J2000). East is at top; the image spans  $7.7 \times 2.4$  arcmin. The top panel shows the  $r$ -band image (units:  $\text{nmgy arcsec}^{-2}$ , square root stretch), and the centre panel shows the inverse variance map (units:  $\text{nmgy}^{-2} \text{arcsec}^2$ , linear stretch). Note the dark vertical stripes in the inverse variance map produced by bad columns, and the square patches due to cosmic ray hits propagating through the masking procedure. The spiral galaxy near the centre of the image is of magnitude  $r = 17.4$ . The single-epoch image is from strip 82S, run 4263, field 310, camcol 1 (acquired on 2003 November 20 at airmass 1.21). The image shown is the fpC image from rerun 40 on the Data Archive Server (units: uncalibrated, linear stretch). The same number of pixels are shown, but note that the single-epoch image is at the native pixel scale (0.396 instead of 0.36 arcsec) and hence shows a slightly larger area.

$r$  band. Objects with multiple peaks are deblended: the deblending algorithm assigns image flux to each peak in the parent object.<sup>12</sup>

<sup>12</sup> Short descriptions of the SDSS deblending can be found in Stoughton et al. (2002, section 4.4.3) and on the SDSS website at <http://www.sdss.org/dr7/algorithms/deblend.html>. A detailed paper describing the deblender is forthcoming (Lupton et al., in preparation).

Once a complete list of deblended peaks (hereafter objects) has been constructed, the properties of each peak are measured. For the purposes of this paper, the most important outputs are the MODELFLUX and MODELFLUX\_IVAR parameters,<sup>13</sup> which are determined from the total flux in the best-fitting (PSF-convolved) galaxy

<sup>13</sup> <http://www.sdss3.org/dr8/algorithms/magnitudes.php>

**Table 2.** Masking radius as a function of apparent stellar magnitude.

Magnitude range	Masking radius (arcsec)
$r, i < 12$	100
$12 < r, i < 13$	70
$13 < r, i < 14$	50
$14 < r, i < 15$	40
$15 < r, i < 16$	30

profile in the  $r$  band (comparing the likelihoods for an exponential and a de Vaucouleurs model), with the amplitude re-fitted separately to each of the other bands. This flux measure approximates the true, total flux in the  $r$  band, and provides a robust colour measurement, which is crucial for photometric estimates of the object redshift distribution.

The final crucial output of PHOTO-FRAMES, for lensing purposes, is a postage stamp image for every unique object detected in the catalogue, except for those objects for which the deblender algorithm failed.

#### 4.8 Lensing catalogue construction

After PHOTO-FRAMES has constructed an object catalogue from the co-added images, we attempt to eliminate spurious detections, stars and galaxies that are unsuitable for shape measurement. Information from the input mask (fpM) files is propagated through to the catalogue, so that objects that incorporate bad pixels identified earlier in the pipeline can be excluded as needed. Due to the nature of the kImages produced by the image co-addition, many of the standard SDSS flags will not be used (e.g., by construction, there are no saturated pixels). As we describe above, masked regions of the kImages are marked as interpolated; objects in the photometric catalogue outputs with these bits set are removed from the catalogue at this stage. Any galaxies on which the deblending algorithm failed are also excluded, as PHOTO-FRAMES will not generate unique postage stamps for these objects.

PHOTO-FRAMES also attempts to classify objects as ‘stars’ or ‘galaxies’ on the basis of the relative fluxes in the PSF and galaxy model fits (Lupton et al., in preparation). Objects that are well described by a PSF are classified as stars; we do not include these objects in the shape catalogue, but set them aside as aids for detecting systematic errors.

To minimize these effects, we also match against a list of all objects classified as stars in the single-epoch SDSS catalogues<sup>14</sup> with apparent magnitudes in the  $i$  or  $r$  band brighter than 15. We remove objects from the catalogue within an angular separation of these bright stars that depends on the stellar apparent magnitude as described in Table 2.

In addition to these basic cuts, we cull the following objects from the lensing catalogue:

- (i) all objects where the model flux or ellipticity moment measurement failed;
- (ii) all objects within 62 pixels of the beginning or end of a frame;
- (iii) all objects detected only in the binned images (BINNED2 or BINNED4);

<sup>14</sup> As our sky coverage is less complete than the single-epoch data, we use the single-epoch catalogues in masking so as to remove objects that are in close proximity to a star that is in one of our masked regions.

(iv) all objects where a bad pixel was close to the object centre (INTERP\_CENTER) in either of the  $r$  or  $i$  bands, or both

(v) all objects that are parents of blends (i.e. measured again in terms of the individual child objects);

(vi) those for which the observed  $r$ -band magnitude is greater than 23.5 or the  $i$ -band magnitude is greater than 22.5.

The magnitude cut was applied to ‘observed’ (at the top of the atmosphere) rather than Galactic extinction-corrected magnitudes. While this leads to a non-uniform galaxy number density, it avoids issues with the limiting- $S/N$  varying with position. Using the Schlegel, Finkbeiner & Davis (1998) dust map, with the standard extinction-to-reddening ratios (Stoughton et al. 2002, table 22), along the occupied 100° length of the stripe, the  $r$ -band extinction  $A_r$  has a mean value of 0.141 and a standard deviation of 0.065. (The  $i$ -band extinction is lower by a factor of 0.76.) A simple test using the COSMOS Mock Catalogue (CMC; Jovel et al. 2009) and a size cut<sup>15</sup> at  $r_{\text{eff}} > 0.47$  arcsec predicts that this standard deviation should result in a  $1\sigma$  variation of  $\pm 3$  per cent in the galaxy density and  $\pm 1$  per cent in the mean redshift  $\langle z \rangle$ . The systematic error introduced by non-uniform depth, which should be second order in the amplitude of variations, is expected to be negligible for the purposes of the SDSS analysis. Note however that this will not be true of future projects.

Many of these cuts are applied in only one band. The result of this process is to produce two separate shape catalogues, one for each of the two shape-measurement bands, so there are a small number of galaxies which appear in only one of the two catalogues.

The SDSS photometric pipeline is known to produce significant sky proximity effects, wherein the photometric properties of objects detected near a bright star are systematically biased. The effect of bright stars on the measured tangential shear of nearby galaxies in single-epoch SDSS data is shown in Fig. 3. Motivated by the scales of the effects seen there, we mask the regions around bright stars with a masking radius that depends on the apparent  $r$ -band (model) magnitude of the stars as given in Table 2.

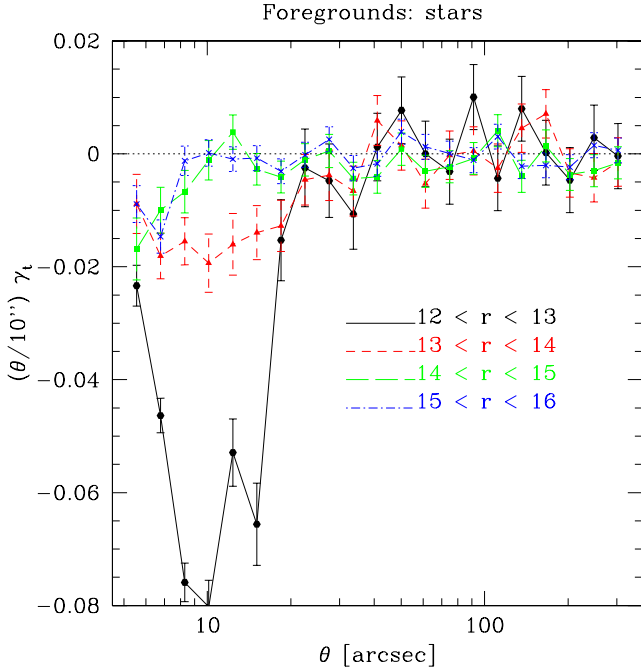
#### 4.9 Shape measurement

Once the final shape catalogue has been constructed, we use the re-Gaussianization shape-measurement method of Hirata & Seljak (2003) to generate an ellipticity measure for each object. The processing code and script are a modification of those used in Mandelbaum et al. (2005). Re-Gaussianization is not an especially modern shape-measurement technique, but we have used it previously on SDSS data, it meets our requirements for shear calibration given the expected statistical power, and we had a well-tested code that interfaced to PHOTO-FRAMES outputs at the time of initiating the cosmic shear project. Therefore, we chose to continue using it for this analysis.

##### 4.9.1 Overview of re-Gaussianization

The re-Gaussianization method was an outgrowth of previous work by Bernstein & Jarvis (2002). They defined the adaptive moments  $\mathbf{M}_l$  of an image  $I$  by finding the Gaussian  $\mathcal{G}[I]$  that minimizes

<sup>15</sup> For an  $r_{\text{eff}}$  of the PSF of 0.67 arcsec and a resolution factor cut at  $R_2 > 0.333$ , we expect the minimum  $r_{\text{eff}}$  of a usable galaxy to be  $0.67\sqrt{0.333/(1-0.333)}$  arcsec. This is of course only a very rough estimate, but this application of the CMC provides a simple and fast way to estimate the impact of marginal changes in survey parameters.



**Figure 3.** Tangential shear  $\gamma_t$  for galaxies as a function of separation  $\theta$  from stars, as measured in the single-epoch SDSS imaging using the shape catalogue from Reyes et al. (2012). The different lines with points show results for bins in  $r$ -band stellar apparent magnitude, as labelled on the plot. The ideal expected value of zero is shown as a dotted horizontal line.

the  $L^2$  norm  $\|I - \mathcal{G}[I]\|$ . A Gaussian has six parameters – an amplitude, two centroids  $\bar{x}_I$  and three components of the symmetric covariance matrix – and the last of these is by definition the  $2 \times 2$  adaptive moment matrix. The ellipticity of the galaxy is defined via its components

$$e_+^{(f)} = \frac{\mathbf{M}_{f,xx} - \mathbf{M}_{f,yy}}{\mathbf{M}_{f,xx} + \mathbf{M}_{f,yy}} \quad (24)$$

and

$$e_\times^{(f)} = \frac{2\mathbf{M}_{f,xy}}{\mathbf{M}_{f,xx} + \mathbf{M}_{f,yy}}. \quad (25)$$

For Gaussian PSFs and galaxies, it is easily seen that the adaptive moment of the intrinsic galaxy image  $f$  can be extracted from that of the observed image via  $\mathbf{M}_f = \mathbf{M}_I - \mathbf{M}_\Gamma$ . If the PSF is both circular and Gaussian (a situation that does not arise in practise), then one can relate the ellipticity of the observed image to that of the true galaxy image via the resolution factor  $R_2$ :

$$e^{(f)} = \frac{e^{(I)}}{R_2} \quad \text{and} \quad R_2 = 1 - \frac{T_\Gamma}{T_I}, \quad (26)$$

where we have used  $T$  to denote the trace of the adaptive moment matrix: e.g.  $T_\Gamma \equiv \mathbf{M}_{\Gamma,xx} + \mathbf{M}_{\Gamma,yy}$ . Re-Gaussianization seeks to apply corrections to equation (26) to correct for the non-Gaussianity of the PSF and the galaxy.<sup>16</sup>

<sup>16</sup> There are also steps in the Hirata & Seljak (2003) code that correct for non-circularity of the PSF. However, since the co-add code has already circularized the PSF, these portions of the code are vestigial and we do not describe them here.

#### 4.9.2 Non-Gaussian galaxies

The shape-measurement procedure described in the previous section is only valid where the PSF itself is Gaussian. One additional step is required to account for the fact that our rounding kernel was chosen so as to produce a PSF that is the sum of two Gaussians. First is the non-Gaussian galaxy correction – i.e. we consider the case of a Gaussian PSF and non-Gaussian galaxy. Appendix C of Bernstein & Jarvis (2002) used a Taylor expansion method to show that if the galaxy is well resolved, then in this case equation (26) could be corrected by using a different formula for the resolution factor,

$$R_2 = 1 - \frac{(1 + \beta_{22}^{(I)}) T_\Gamma}{(1 - \beta_{22}^{(I)}) T_I}, \quad (27)$$

where  $\beta_{22}^{(I)}$  is the radial fourth moment,

$$\beta_{22}^{(I)} = \frac{\int (\rho^4 - 4\rho^2 + 2) I(\mathbf{x}) \mathcal{G}[I](\mathbf{x}) d^2\mathbf{x}}{2 \int I(\mathbf{x}) \mathcal{G}[I](\mathbf{x}) d^2\mathbf{x}}, \quad (28)$$

where  $\mathcal{G}[I]$  is the adaptive Gaussian and the rescaled radius  $\rho$  is given by

$$\rho \equiv \sqrt{(\mathbf{x} - \bar{\mathbf{x}}_I) \cdot \mathbf{M}_I^{-1} (\mathbf{x} - \bar{\mathbf{x}}_I)}. \quad (29)$$

This is equivalent to an elliptical version of the  $n = 4, m = 0$  polar shapelet (Refregier 2003; Refregier & Bacon 2003), and we have  $\beta_{22}^{(I)} = 0$  for a Gaussian galaxy (in practice usually  $\beta_{22}^{(I)} > 0$ ).

#### 4.9.3 Non-Gaussian PSF

The shape-measurement procedure described in the previous section is only valid where the PSF itself is Gaussian. One additional step is required to account for the fact that our rounding kernel was chosen so as to produce a PSF that is the sum of two Gaussians. We start by constructing a Gaussian approximation  $G_1$  to the PSF  $G$ ,

$$\Gamma(\mathbf{x}) \approx G_1(\mathbf{x}) = \frac{1}{2\pi\sqrt{\det \mathbf{M}_{G_1}}} \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{M}_{G_1}^{-1} \mathbf{x}\right). \quad (30)$$

The choice  $G_1$  is chosen according to the adaptive moments of  $\Gamma$ . The function  $G_1$  is determined from the centroid and covariance, but the amplitude in equation (30) is chosen to normalize the Gaussian  $G_1$  to integrate to unity.<sup>17</sup>

We may then define the residual function  $\epsilon(\mathbf{x}) = \Gamma(\mathbf{x}) - G_1(\mathbf{x})$ . It follows that the measured image intensity will satisfy  $I = \Gamma \otimes f = G_1 \otimes f + \epsilon \otimes f$ , where  $\otimes$  represents convolution. This can be rearranged to yield

$$G_1 \otimes f = I - \epsilon \otimes f. \quad (31)$$

This equation thus allows us to determine the Gaussian-convolved intrinsic galaxy image  $I$ , if we know  $f$ . At first glance, this does not appear helpful, since if we knew  $f$  it would be trivial to determine  $\Gamma \otimes f$ . However,  $f$  appears in this equation multiplied by (technically, convolved with) a small correction  $\epsilon$ , so equation (31) may be reasonably accurate even if we use an approximate form for  $f$ . The simplest approach is to approximate  $f$  as a Gaussian with

<sup>17</sup> The reason for doing this is that while this increases the overall power  $\int \epsilon^2$  of the residual function, it yields  $\int \epsilon = 0$ , which ensures that for well-resolved objects (i.e. objects for which the PSF is essentially a  $\delta$ -function), the ‘correction’  $\epsilon \otimes f_0$  applied by equation (33) does not corrupt the image  $I$ .



**Table 3.** Parameters of the shape catalogue.

Parameter	Value		Units
	<i>r</i> band	<i>i</i> band	
Total number of source galaxies	1328 885		
Number of sources per band	1067 031	1251 285	
Effective number of sources downweighted by noise, $N_{\text{eff}} = \sum_i \varpi_i$	882 345	1065 807	
Median magnitude	21.5	20.9	mag AB
Median resolution factor $R_2$	0.55	0.53	
rms measured ellipticity per component (noise not subtracted)	0.48	0.47	

covariance:

$$f_0 = \frac{1}{2\pi\sqrt{\det \mathbf{M}_f^{(0)}}} \exp\left(-\frac{1}{2}\mathbf{x}^T [\mathbf{M}_f^{(0)}]^{-1} \mathbf{x}\right), \text{ with}$$

$$\mathbf{M}_f^{(0)} = \mathbf{M}_I - \mathbf{M}_\Gamma, \quad (32)$$

where  $\mathbf{M}_I$  and  $\mathbf{M}_\Gamma$  are the adaptive covariances of the measured object and PSF, respectively. Then we can define

$$I' \equiv I - \epsilon \otimes f_0 (\approx \Gamma \otimes f). \quad (33)$$

The adaptive moments of  $I'$  can then be computed, and the PSF correction of equation (29) can then be applied to recover the intrinsic ellipticity  $e^{(f)}$ .

Simple simulations with (noise-free) toy galaxy profiles indicate that this method has shear calibration errors at the level of a few per cent depending on the galaxy profile, with the worst performance for de Vaucouleurs profiles at low resolution and high ellipticity (Mandelbaum et al. 2005, fig. 5). Moreover, simulations of SDSS data based on real galaxy profiles from COSMOS, single-epoch SDSS PSFs and realistic noise levels show that the shear calibration biases are not markedly different under more realistic conditions (Mandelbaum et al. 2012). An investigation of the shear calibration bias for the SDSS cosmic shear sample is presented in Huff et al. (2014), the second paper in this series (hereafter Paper II).

To select the galaxy sample used for the final analysis, we impose a resolution factor cut at  $R_2 > 0.333$  in both *r* and *i* (we will justify this choice in Section 6.3 based on our desire to minimize additive PSF systematics). The parameters of the final shape catalogue are shown in Table 3, and the survey geometry can be found in Fig. 4. The apparent magnitude distribution in each band is shown in Fig. 5. We show a comparison with the single-epoch photometry for a representative subsample of galaxies in Fig. 6.

## 5 CORRELATION FUNCTION ESTIMATION

As stated previously, the primary systematic error of concern in this paper is additive shear systematics, due to PSF ellipticity leaking into the galaxy shapes even after the PSF correction is carried out. This concern will drive our choice of diagnostics to use on the shape catalogues. There are several possible choices for diagnostics that we could use.

(i) One-point statistics of the star and galaxy shapes. For example, we calculate the mean stellar and galaxy ellipticities in bins of some chosen size and look for deviations from zero, including coherent patterns. We use this diagnostic in Section 6.1.

(ii) The tangential shear as a function of scale around random points (e.g. Mandelbaum et al. 2005). If there is some additive systematic shear, then on scales that are such that we start losing lens–source pairs off the survey edge, it will show up as a non-zero tangential shear. However, this test alone does not tell us much about

the correlations between systematic shears at different points, and therefore we ignore it in favour of more informative tests.

(iii) Cross-correlations between the stellar shapes and galaxy shapes, as a function of separation  $\theta$ . These correlation functions tell us not only about the amplitude of any systematic shear, but also about the characteristic scales that are affected by it. This section will describe our methodology for calculating these correlation functions.

(iv) The *B*-mode shear, which should be zero due to gravitational lensing. While this test is an important one as it can signal a variety of problems with PSF correction, it is not strictly a measure of additive shear systematics. Thus, we leave this test for Paper II, which presents the cosmic shear analysis.

### 5.1 The estimator and weighting

In order to compute the star–galaxy cross-correlations, we employ a direct pair-count correlation function code. It is slow ( $\sim 3$  h for  $2 \times 10^6$  galaxies on a modern laptop) but robust and well adapted to the Stripe 82 survey geometry. The code sorts the galaxies in order of increasing right ascension  $\alpha$ ; stars and galaxies are assigned to the range  $-60^\circ < \alpha < +60^\circ$  to avoid unphysical edge effects near  $\alpha = 0$ . It then loops over all pairs with  $|\alpha_1 - \alpha_2| < \theta_{\text{max}}$ . The usual ellipticity correlation functions can be computed via summation over galaxies *i* and stars *j*, e.g.

$$\xi_{11,\text{PSF}}(\theta) = \frac{\sum_{\alpha\beta} w_i e_{\alpha 1} \mathbf{M}_E \mathbf{1}_\beta}{\sum_{\alpha\beta} w_\alpha} \quad (34)$$

and similarly for  $\xi_{22,\text{PSF}}$ . Here  $e_{i1}$  is the PSF-corrected galaxy  $e_1$  for galaxy index *i*, and  $\mathbf{M}_E \mathbf{1}_j$  is the stellar  $e_1$  derived from the adaptive moments described in Section 4.9. The sum is over pairs with separation in the relevant  $\theta$  bin, and we weight each pair according only to the weight associated with the galaxy in each pair:

$$w_i = \frac{1}{0.37^2 + \sigma_e^2}. \quad (35)$$

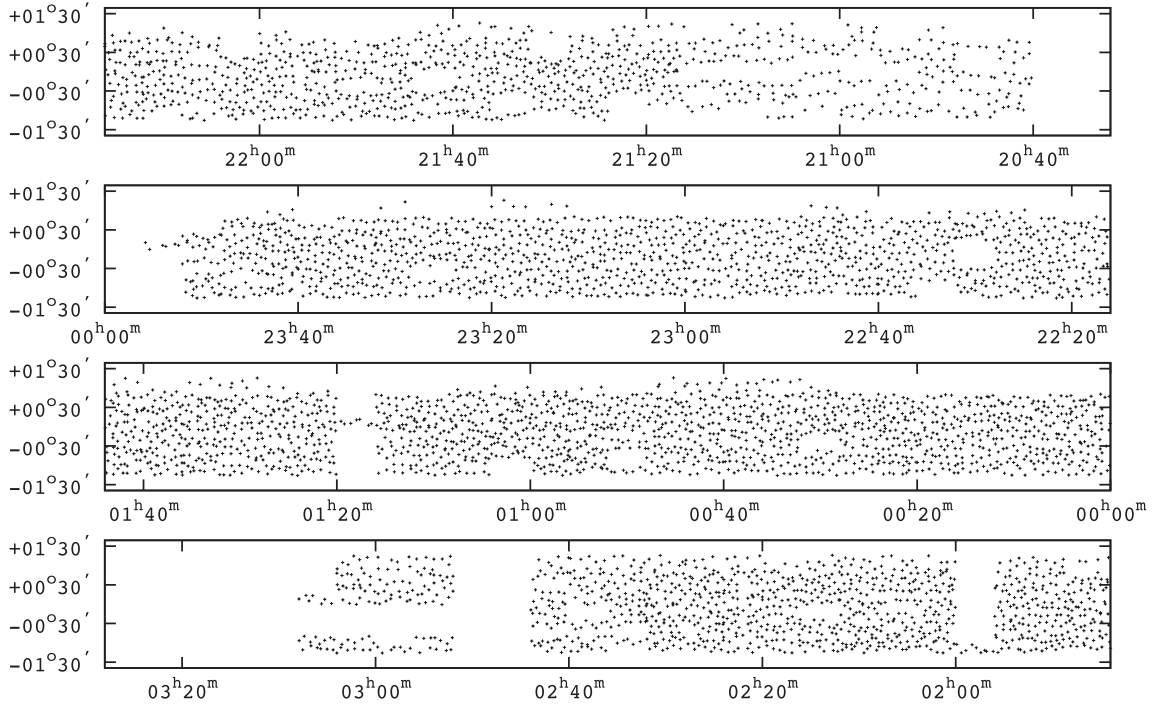
Following Reyes et al. (2012), we have for weighting purposes adopted an intrinsic shape noise  $e_{\text{rms}}$  per component of 0.37. The weight of a galaxy relative to a galaxy with perfectly measured shape is

$$\varpi_i = \frac{w_i}{w(\sigma_e = 0)} = \frac{1}{1 + \sigma_e^2/0.37^2}. \quad (36)$$

Since the imaging is taken in drift-scan mode, which introduces a potential preferred direction for PSF distortions, we compute our diagnostic correlations between the components aligned along ( $-e_1$  and  $-\mathbf{M}_E \mathbf{1}$ ) and at  $45^\circ$  to ( $e_2$  and  $\mathbf{M}_E \mathbf{e}_2$ ) the scan direction.

The code works on a flat sky, i.e. equatorial coordinates ( $\alpha$ ,  $\delta$ ) are approximated as Cartesian coordinates. This is appropriate in





**Figure 4.** The angular distribution (in J2000 right ascension and declination) of the *i*-band galaxy catalogue. A subsample of every 250th galaxy is shown. The *r*-band sample is identical except for the missing range of  $-00^{\circ}48' < \text{Dec.} < -00^{\circ}24'$ . Note the complex survey geometry. Coverage gaps at  $\text{Dec.} > 0.8$  are primarily due to the severe PSF quality cuts made during the image co-addition step.

the range considered,  $|\delta| < 1.274$ , where the maximum distance distortions are  $\frac{1}{2}\delta_{\text{max}}^2 = 2.5 \times 10^{-4}$ .

All of our shape correlations are computed over the range  $1 < \theta < 120$  arcmin, evenly spaced in  $\log \theta$ .

## 5.2 Statistical errors

The direct pair-count correlation function code can directly compute the Poisson error bars, i.e. the error bars neglecting the correlations in  $e_{i\alpha} \mathbf{M} \mathbf{E} \alpha_j$  between different pairs. This estimate of the error bar is

$$\sigma^2[\xi_{++}(\theta)] = \frac{\sum_i w_i^2 |e_i|^2 |\mathbf{M} \mathbf{E} j|^2}{2 \left( \sum_i w_i \right)^2}. \quad (37)$$

Equivalently, this is the variance in the correlation function that one would estimate if one randomly re-oriented all of the galaxies. As the star–galaxy correlations described here are approximate indicators of the amplitude of the additive PSF shear, and not precision estimates for use in a cosmic shear analysis, we will not attempt to infer the covariance matrix for the full diagonal star–galaxy cross-correlation functions.

## 6 DIAGNOSTICS

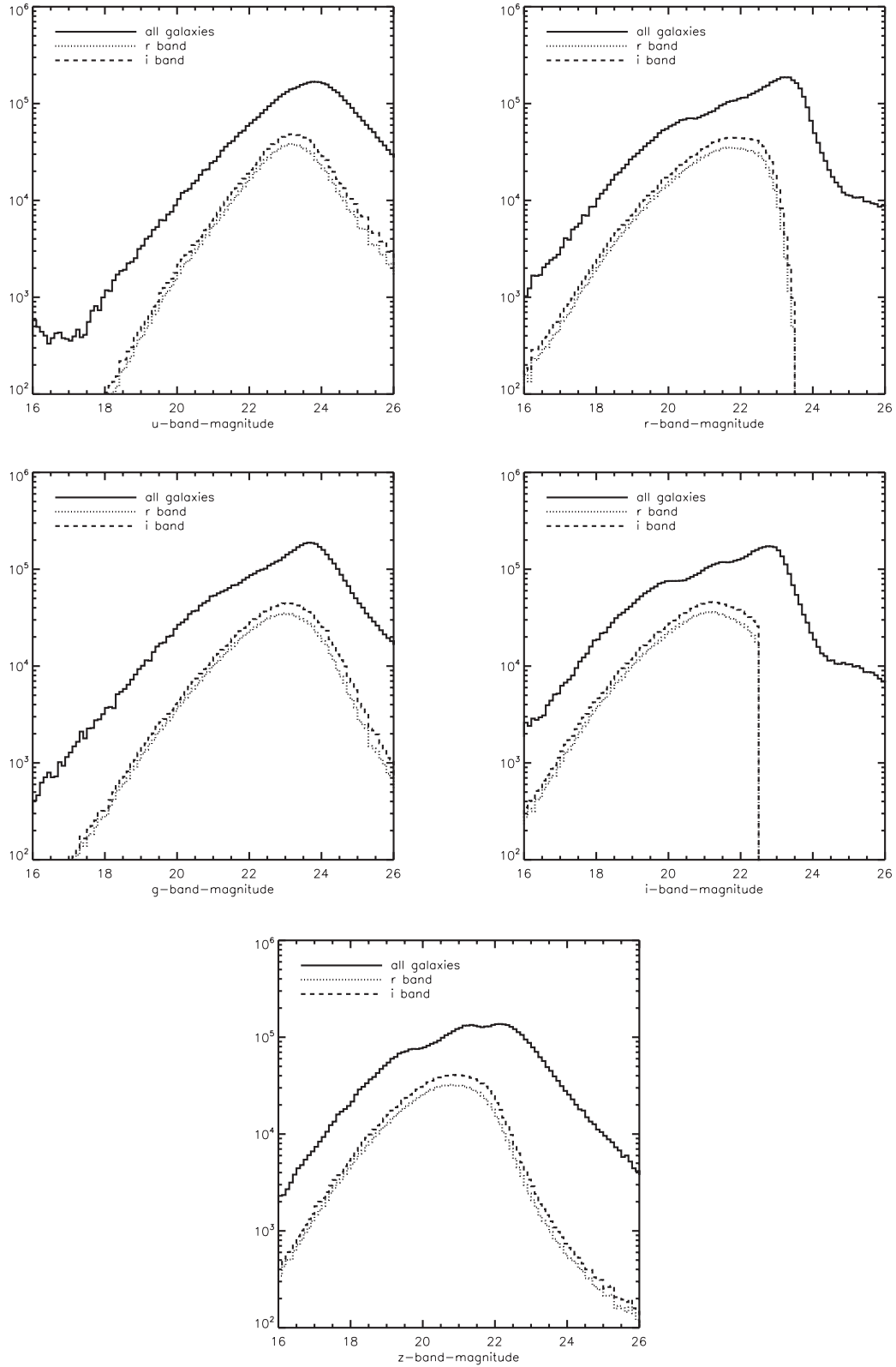
Here we present our two main systematics tests described in Section 5, namely the one-point statistics of the stellar and galaxy ellipticities and the star–galaxy shape cross-correlations. In order to do this calculation, we must define a star catalogue, which relies on the PHOTO star–galaxy separation. The colours of the objects selected as stars by PHOTO are shown in Fig. 7. As shown, they agree with previous determinations of the colours of the stellar locus, e.g. from Richards et al. (2002).

### 6.1 Average shapes

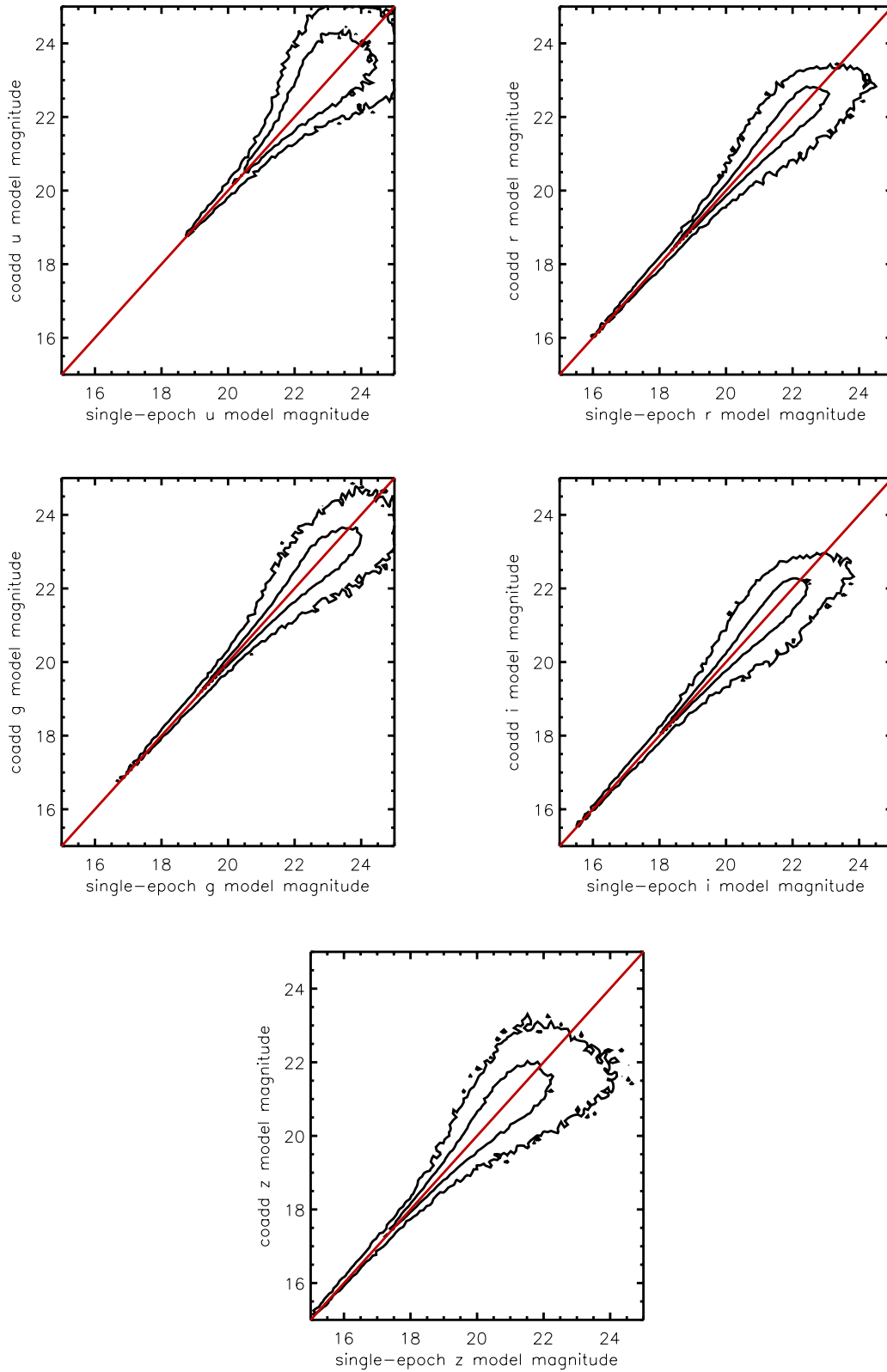
We first estimate the influence of residual PSF ellipticities on the galaxy shapes by mapping the stellar shape field.

We computed a set of star shapes binned by right ascension and declination. The stars were chosen to be moderately faint,  $19.5 < r < 21.5$ , such that they were *not* used to estimate the PSF model in the single-epoch images that was used to construct the rounding kernel applied to each single-epoch image. Fig. 8 shows the results: the mean stellar ellipticities are usually small, of the order of  $10^{-3}$ , but in the *r* band in a particular declination range covered by camcol 2, the shapes are systematically elongated in the scan direction by  $-e_1 = 0.005$ . We find no significant changes in the amplitude of this artefact when splitting the stellar populations by colour ( $r - i < \text{or} > 0.3$ ) or by apparent magnitude ( $r < \text{or} > 20.5$ ). We did not definitively determine the source of this elongation, but we have confirmed that it appears in the single-epoch SDSS imaging (including the galaxy shape catalogues from Mandelbaum et al. 2005; Reyes et al. 2012), so is not merely an artefact of the co-addition and catalogue-making process of this work.<sup>18</sup> There is no counterpart feature in the *i* band. Given the fact that this feature may plausibly arise due to problems with the single-epoch PSF model used to determine the proper convolution kernel to achieve the desired co-add PSF, we exclude all *r*-band galaxy data in camera column 2 from the cosmic shear analysis.

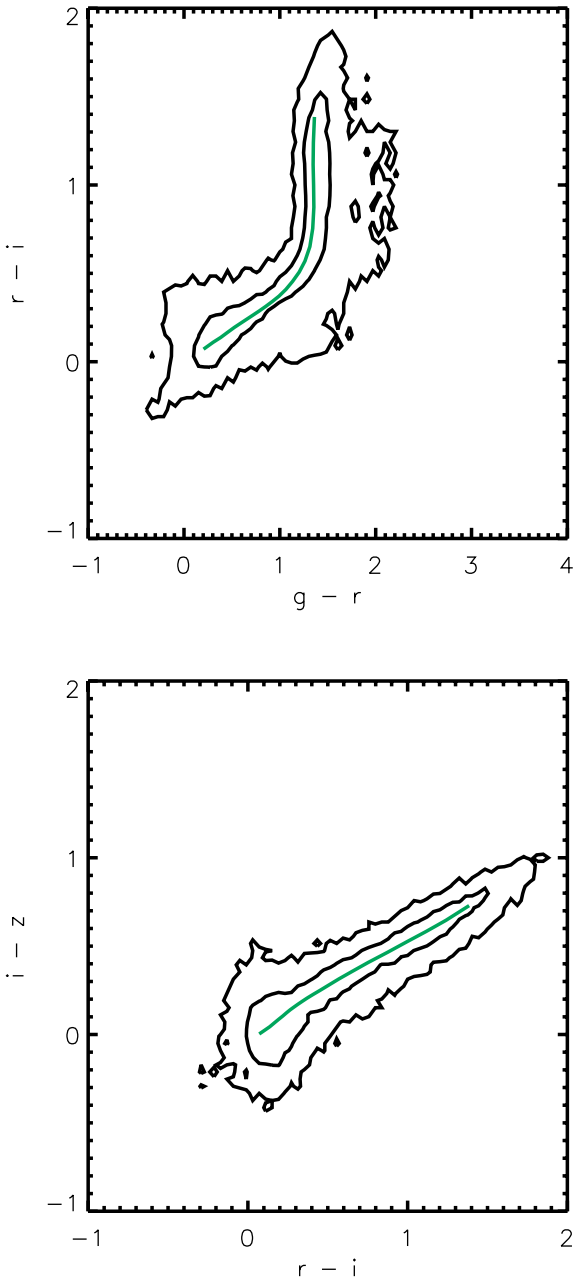
<sup>18</sup> One possible explanation is incorrect non-linearity corrections for the *r*-band camcol 2 CCD. The stars used to construct the PSF model are sufficiently bright that they require non-linearity corrections, but the stars used for our tests here do not. Therefore, if the non-linearity correction is wrong for that CCD, it could affect the PSF model for that CCD alone.



**Figure 5.** The distribution of observed (not corrected for Galactic extinction) apparent galaxy model magnitudes in the  $u$ ,  $g$ ,  $r$ ,  $i$  and  $z$  bands (top-left, middle-left, top-right, middle-right and bottom panels). In all cases, the solid line shows the apparent magnitudes for all unique extended objects; the dotted and dashed show the  $r$ - and  $i$ -band lensing catalogues, respectively.



**Figure 6.** The comparison of the observed (not corrected for Galactic extinction) model magnitudes of galaxies in the co-add lensing catalogue with magnitudes for the same objects in the best run at that position in the single-epoch imaging. Contours are 68 and 95 per cent of the total matches. The asymmetry around the 1:1 line at faint magnitudes is due to the flux limit in the single-epoch images.



**Figure 7.** Density contour plots in colour–colour space for objects identified as stars using PHOTO’s star–galaxy separation based on the concentration of the light profile; the contours containing 68 and 95 per cent of the density are shown. The stellar locus from Richards et al. (2002) is shown as a solid line. This plot includes correction for Galactic extinction, for fair comparison with previous results.

## 6.2 Star–galaxy cross-correlation

Our primary tasks in producing a shear measurement are to demonstrate that the additive systematic shear is below the target threshold set above (Section 2), and that our shape-measurement method allows us to correctly translate the measured shapes into shears with sufficient accuracy (a task that we will handle in more detail in Paper II). We perform two classes of such tests. The first, using stars in the co-added images that were too faint to use in the original PSF determination (Section 6.2.1), is directly related to the correlation function of the residual PSF ellipticity field in the co-added images.

This test is sensitive to e.g. errors associated with interpolation of the PSF from the positions of the brighter PSF stars. The second test (Section 6.2.2) measures the correlations of galaxy ellipticity with the uncorrected stars in the original images. It is sensitive to any ‘leakage’ of the original ellipticity field into the galaxy ellipticity catalogue used for the cosmic shear analysis. This correlation would not exist under ideal circumstances but could be present if e.g. the PSF ellipticity exhibited a flux dependence (having a similar spatial pattern but a different amplitude for stars and galaxies).

### 6.2.1 Tests with stars in the co-added images

In order to test for residual additive shear systematics, we calculate the cross-correlation between the measured shapes of the stars and those of the galaxies in our sample. Any remaining contribution to the inferred shear field of the galaxies that is sourced by the PSF will produce a non-zero cross-correlation. It is important to note that this measurement is performed using measurements of the images of stars *not* used to construct the model PSF; the shape measurements of these objects are not in any sense corrected, and do not incorporate knowledge of the PSF model. As a result, the star–galaxy shape correlations are diagnostic of any spatially varying PSF modelling errors. Constant multiplicative errors due, for example, to finite-pixel effects, noise bias or similar problems will be handled in the shear calibration step in Paper II.

We estimate the star–star and star–galaxy cross-correlations as in equation (34) for all star–galaxy pairs within and between the  $r$  and  $i$  bands. The results for the star–galaxy correlations are shown in Fig. 9. For the systematic error diagnostics considered here, we are primarily interested in computing the cross-correlation between resolved galaxies and unresolved point sources.

### 6.2.2 Tests with uncorrected stars in the single-epoch images

The second class of tests is designed to test for any residual signal in the galaxy ellipticity catalogue that correlates with the original PSF ellipticity field. The key test in this case is to compute the correlation function of *uncorrected* stars in the single-epoch images, and the final *corrected* galaxy catalogue. This test (unlike the previous one) was performed using the final galaxy catalogue including the declination-dependent subtraction from Paper II – i.e. it was performed on the same version of the galaxy catalogue used for the cosmic shear analysis.

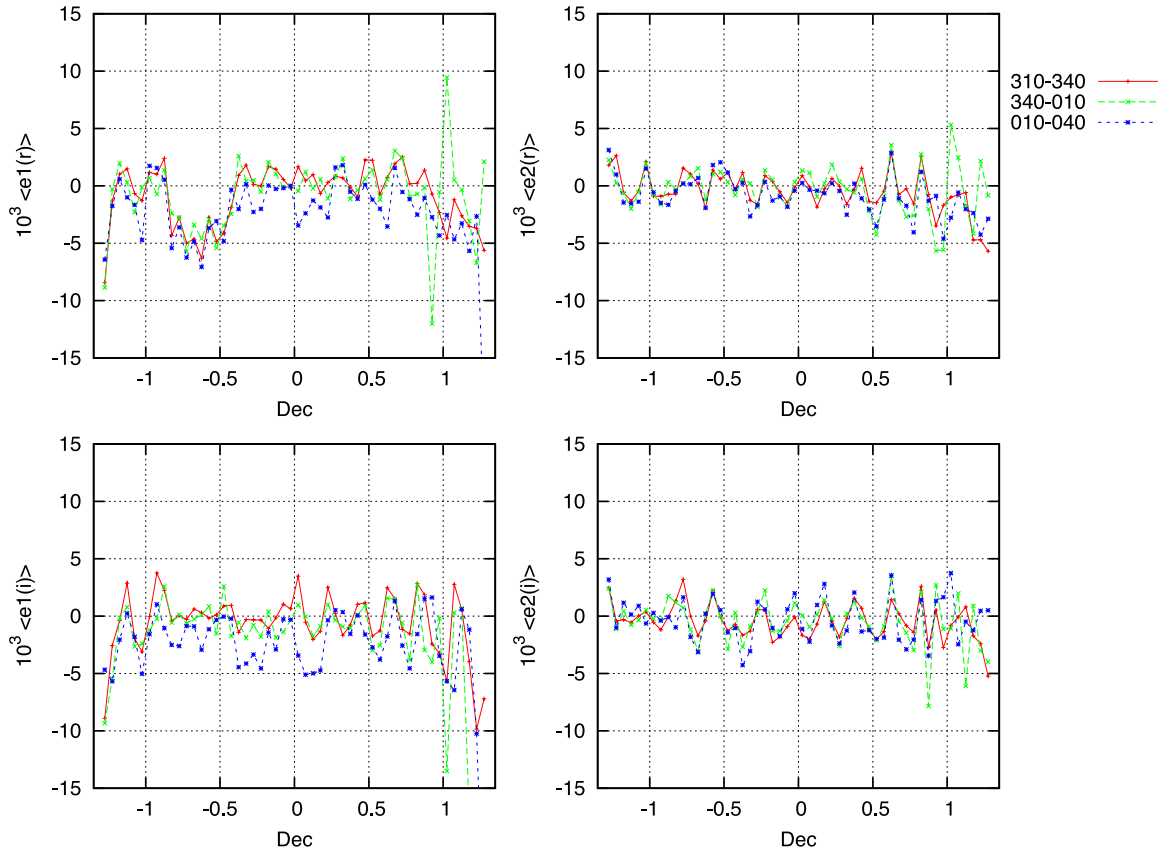
To construct the star catalogue from single-epoch SDSS imaging, we first matched those stars used for the tests in Section 6.2.1 to the catalogues of single-epoch SDSS Stripe 82 imaging, using a 1 arcsec matching radius. For each match, we computed the mean ellipticity over all those runs that were used in the co-adds and in which the star was detected.

The correlation function  $\xi_{sg}(\theta)$  was computed for both  $++$  and  $\times\times$  components, and in both the  $r$  and  $i$  bands. The results are shown in Fig. 10.

The error bars on the star–galaxy correlation function were computed assuming uncorrelated ellipticities on the galaxies. That is, for each galaxy  $i$  and ellipticity component  $\alpha \in \{1, 2\}$ , and each correlation function bin, we compute the partial derivative  $\partial \xi_{sg}(\theta) / \partial e_{i\alpha}$ . We then approximate the covariance matrix as

$$\text{Cov}[\xi_{sg}(\theta), \xi_{sg}(\theta')] = \sum_{i\alpha} \frac{1}{2} e_i^2 \frac{\partial \xi_{sg}(\theta)}{\partial e_{i\alpha}} \frac{\partial \xi_{sg}(\theta')}{\partial e_{i\alpha}}; \quad (38)$$





**Figure 8.** The mean ellipticities of stars in the  $r$  band as a function of declination for different ranges of right ascension, as indicated at the upper right. The top panels show the  $r$  band and the bottom panels show the  $i$  band, while the left- and right-hand panels show different ellipticity components. This was computed using a version of the star catalogue prior to final cuts. Note the spurious effect in camcol 2  $r$  band in the  $e_1$  component (declinations  $-0^{\circ}8$  to  $-0^{\circ}4$ ). The apparent magnitude range for this plot was  $19.5 < r < 21.5$ .

the  $\frac{1}{2}$  takes into account that the squared galaxy ellipticity  $e_i^2$  contains two components. This analytic procedure is equivalent to the covariance matrix that would be obtained by randomizing the orientations of the galaxies. Since the galaxy ellipticity power spectrum in SDSS is dominated by shot noise (rather than true lensing), this is expected to be a good approximation to the true covariance matrix.<sup>19</sup> The diagonal entries in the covariance matrix are plotted in Fig. 10; the full covariance matrices give  $\chi^2$  for the star–galaxy correlation function compared with zero of 13.2 ( $rr++$ ), 8.2 ( $rr\times\times$ ), 8.3 ( $ii++$ ) and 5.6 ( $ii\times\times$ ), all for 10 degrees of freedom.

Also shown in Fig. 10 are the uncorrected star–star ellipticity correlations. These are useful because if the stellar ellipticity field leaks into the galaxy ellipticity field, such that the galaxy ellipticity has a spurious component  $\Delta e_g = C e_s$  (where  $e_s$  is the stellar ellipticity), then we would expect to have a spurious galaxy–galaxy correlation given by  $\Delta \xi_{gg}(\theta) = C \xi_{sg}(\theta)$  and  $\xi_{sg}(\theta) = C \xi_{ss}(\theta)$ . Combining these, we find that the implied spurious galaxy–galaxy correlation is  $\Delta \xi_{gg}(\theta) = \xi_{sg}^2(\theta)/\xi_{ss}(\theta)$ . This contribution is shown in the bottom part of Fig. 10 along with the  $1\sigma$  error on the galaxy correlation function. It is seen that at all scales,  $\xi_{sg}^2(\theta)/\xi_{ss}(\theta)$  is far below the statistical errors. It should however be kept in mind that this test

is only sensitive to systematics that correlate with the uncorrected ellipticities.

### 6.3 Resolution cuts

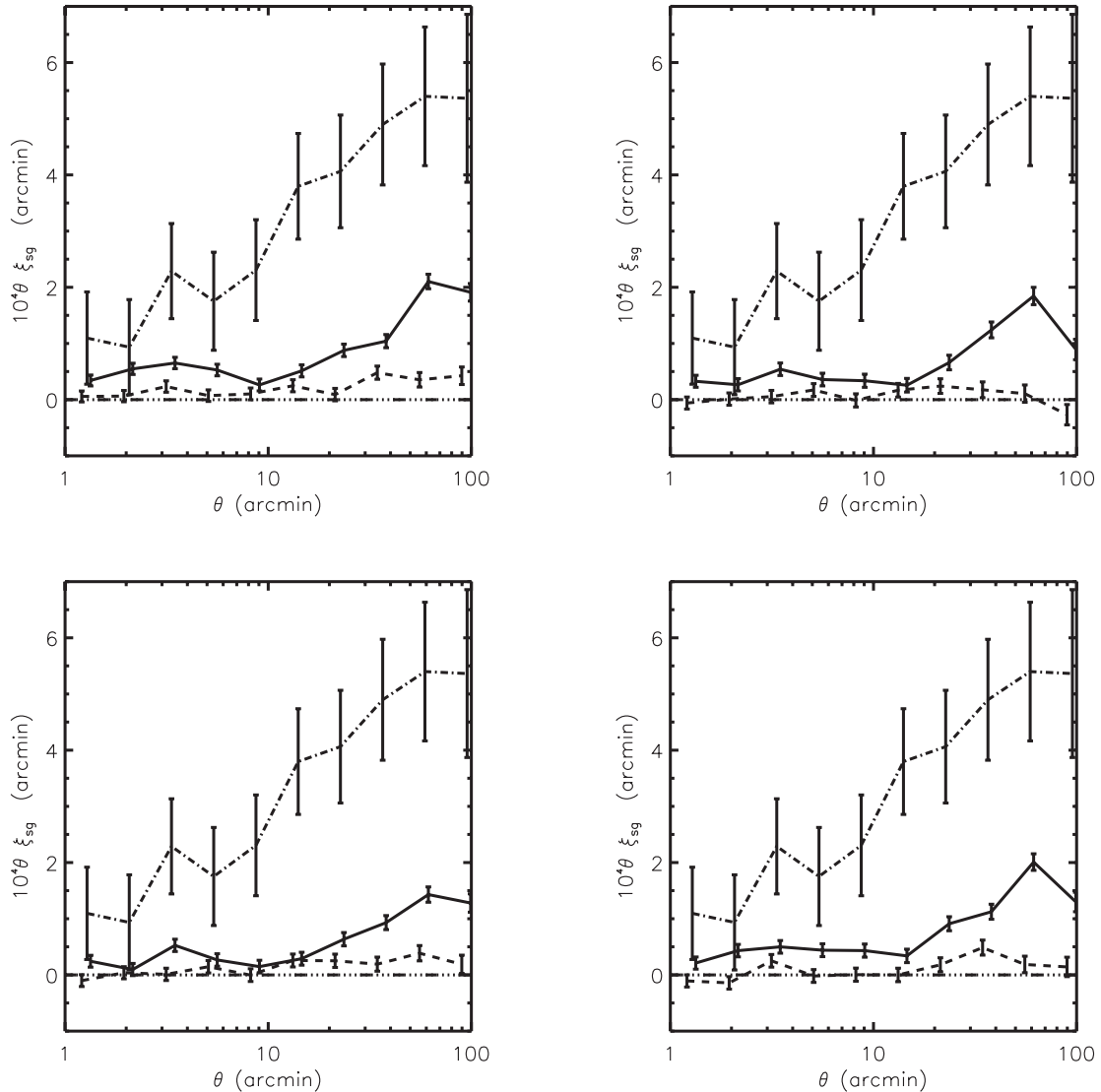
Due to the PSF dilution correction applied to all galaxy shapes in Section 4.9, noisy measurements of poorly resolved galaxies can significantly amplify any residual additive shear systematics not corrected for in the rounding kernel process. To assess the effects of a resolution cut, we compute the star–galaxy cross-correlations in each band for  $R_2 > 0.25$ ,  $> 0.333$  and  $> 0.4$ . Adopting the second of these of these thresholds appears to be sufficient to minimize the amplitude of the star–galaxy shape correlation signal. As a result, we adopt a cut of  $R_2 > 0.333$  for both the  $i$ - and  $r$ -band galaxy catalogues. This resolution cut corresponds to galaxies with a typical half-light radius (as determined from the SDSS model fits) of 0.7 arcsec. Any potential selection bias resulting from resolution cuts will be dealt with in Paper II, when we derive the empirical shear calibration.

### 6.4 Star–galaxy separation

#### 6.4.1 Contamination of star sample by galaxies

A non-zero amplitude of  $\xi_{sg}$  can also be produced by imperfect star–galaxy separation. Poorly resolved galaxies masquerading as stars sample both the PSF- and cosmic shear-sourced shape fields. If the

<sup>19</sup> This is of course a shortcut that will not be applicable to experiments with higher source density: these must compute the covariance matrix via many simulated realizations of the galaxy ellipticity field.



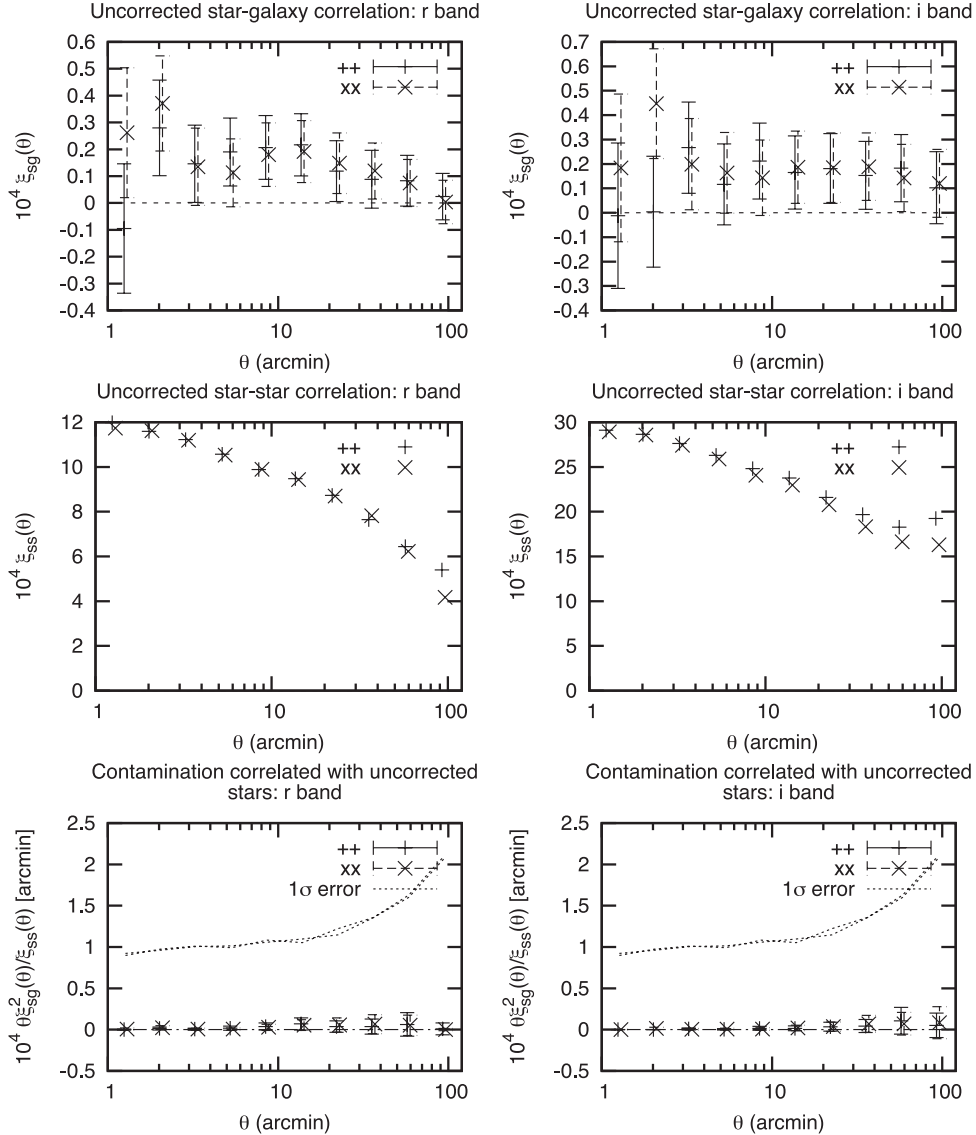
**Figure 9.** The cross-correlation of star shapes with galaxy shapes, for the following pairs of bands:  $(i, i)$  in the upper-left,  $(r, r)$  in the upper-right,  $(r, i)$  in the bottom-left and  $(i, r)$  in the bottom-right panels. All results are shown as  $10^4 \xi$ . The  $\langle e_1 e_1 \rangle$  correlation is the solid line, while the  $\langle e_2 e_2 \rangle$  correlation is the dashed line. The dot-dashed line shows the expected cosmic shear ( $e_+ e_+$ ) shape–shape correlation for a survey of this depth and size, with shot-noise errors. The triple dot-dashed line shows the ideal value of zero for the star–galaxy correlations. Statistics shown are for stars with apparent  $i$ - and  $r$ -band magnitudes between 19.5 and 21.5.

fraction of stars that are actually mistakenly classified as galaxies is  $f_{\text{gal}}$ , then the measured  $\xi_{\text{sg}}$  will include a contribution proportional to  $f_{\text{gal}} \xi_{\gamma}$ . As the ellipticity of nearly unresolved galaxies will be diluted by PSF convolution, this represents an upper limit to the level of star–galaxy correlation that can be introduced via imperfect star–galaxy separation.

The PHOTO-FRAMES pipeline classifies an object as a star or a galaxy on the basis of the relative fluxes of PSF and galaxy model fits to the object’s surface brightness profile. We have already confirmed that we get a reasonable stellar locus from this determination, compared with that from single-epoch imaging (Fig. 7). As another check on this scheme, we have defined a sample of stars for which aperture-matched United Kingdom Infrared Telescope (UKIRT) Infrared Deep Sky Survey (UKIDSS) colours are available. The UKIDSS project is defined in Lawrence et al. (2007). UKIDSS uses the UKIRT Wide Field Camera (Casali et al. 2007). The photometric system is described in Hewett et al. (2006), and the calibration is

described in Hodgkin et al. (2009). The pipeline processing and science archive are described in Hambly et al. (2008). Stars and galaxies separate fairly cleanly in  $j - k$ ,  $r - i$  colour space (e.g. Baldry et al. 2010), so we attempt to use a matched catalogue from Bundy et al. (2012) to put some limits on galactic contamination of the stellar sample (see Fig. 11). This constraint on  $f_{\text{gal}}$  will give us our upper limit  $f_{\text{gal}} \xi_{\gamma}$  on the  $\xi_{\text{sg}}$  due to contamination of the star sample by galaxies.

We match the objects classified as stars in both bands from our co-add to UKIDSS objects with valid  $j - k$  colours; objects with angular separations between the two catalogues less than 1 arcsec are identified. We find 93 753 such stars (as classified by PHOTO). Of these, 11 331, or 12 per cent, have  $j - k$ ,  $r - i$  colours inconsistent with the stellar population. The UKIDSS matches are shallower than the rest of the catalogue in the  $i$  band, but of comparable depth in the  $r$  band. Only 16 per cent of our stars have UKIDSS matches in either band, however, so



**Figure 10.** Tests from the correlation of uncorrected star ellipticities with galaxy ellipticities. Each row has both *r*-band (left) and *i*-band (right) results. Top row: the raw ++ and  $\times\times$  star-galaxy ellipticity correlations. Middle row: the uncorrected star-star ellipticity correlations. Bottom row: the contamination to the galaxy ellipticity correlation function due to components that trace the stellar ellipticity field, computed via  $\xi_{sg}^2(\theta)/\xi_{ss}(\theta)$ . Also shown are the  $1\sigma$  expected errors on the cosmic shear measurement in each of the 10 radial bins (dashed lines). The vertical axis has been scaled by  $\theta$  so that the latter curve can be displayed clearly.

the contamination fraction is not well constrained in the entire star sample.

If, however, this fraction is representative of the galaxy contamination in the entire stellar catalogue, then for an unresolved population with a typical resolution just below our resolution cut, that level of contamination would explain a substantial fraction of the residual PSF systematic amplitude that we see.

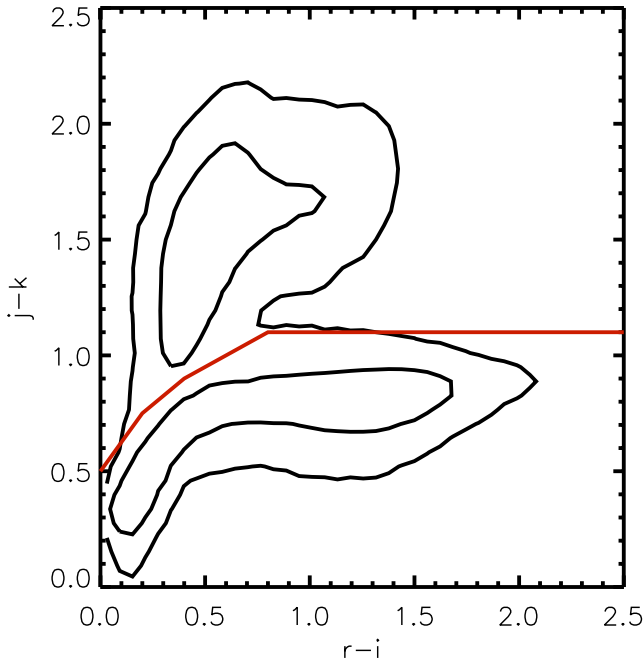
As a test for this, we compute the star-galaxy shape correlation using only those objects identified as stars in the manner described above. The results are shown in Fig. 12. As shown, for this population, the amplitude of the star-galaxy correlation is significantly reduced below the star-galaxy correlations. This is suggestive that some of the star-galaxy signal may arise from galaxy contamination of the star sample. However, because the UKIDSS data do not cover the entire footprint of Stripe 82, this test is not conclusive.

After all of the above cuts have been applied, the final shape catalogue consists of 1067 031 *r*-band and 1251 285 *i*-band shape measurements, over an effective area of 140 and 168 square degrees, respectively.

#### 6.4.2 Contamination of galaxy sample by stars

The other type of contamination, that of the galaxy sample by stars, will tend to dilute lensing statistics measured using our catalogue. Because we wish to understand the contamination in a representative sample of our galaxies (not just the ones bright enough to have a match in the UKIDSS catalogue), we use a different strategy to estimate this type of contamination.

The targeting photometry used for the Deep Extragalactic Evolutionary Probe (DEEP2) survey comes from the CFHT, and in the



**Figure 11.** The cut in  $(j-k, r-i)$  space, defined using extinction-corrected magnitudes, that was used to separate stars from galaxies using the UKIDSS data. Objects below the curve (i.e. blue in  $j-k$ ) are colour-classified as stars, while those above the curve are colour-classified as galaxies.

two DEEP2 fields on Stripe 82, the typical seeing is 0.7–0.8 arcsec, nearly a factor of 2 better than in our co-adds (Coil et al. 2004). The catalogues from this imaging were publicly released<sup>20</sup> as part of DEEP2 Data Release 1. We use the star versus galaxy classification for galaxies in the co-adds that have detections in the DEEP2 targeting photometry as a way to estimate the stellar contamination in our catalogues in those fields.

We first match our galaxy catalogue against the DEEP2 targeting photometry, finding matches for 96 per cent of our galaxies. We then eliminate those that are marked as bad data or saturated in the DEEP2 catalogues. For the remaining objects, the star/galaxy separation works as follows (see Coil et al. 2004 for more details): clearly extended objects are flagged as such, and we consider those as secure galaxy detections. Compact objects are assigned a quantity ‘*pgal*’ in the range [0, 1], representing the probability that the object is a galaxy based on its colour and magnitude. To estimate the total stellar contamination, we sum the values of  $(1 - \text{pgal})$  for all of our galaxies that matched against compact objects, and compare that to the total number of compact and extended matches. The result is a stellar contamination of 1.7 per cent for both the *r*- and the *i*-band lensing catalogues.

One issue in ‘calibrating’ the stellar contamination analysis is that the stellar density varies along the stripe. The Galactic latitudes of the two DEEP2 fields are  $-54^\circ$  and  $-56^\circ$ , but the ‘start’ of our stripe (RA  $310^\circ$ , Dec.  $0^\circ$ ) is at  $l = 46^\circ$ ,  $b = -24^\circ$ . These low-RA fields are not only at lower Galactic latitude, but also look inwards towards the bulge. As a simple test for this, we compared the ratio of the  $19.5 < i < 21.5$  star density to the source galaxy density in the DEEP2 fields (0.16) versus the stripe as a whole (0.46). The ratio of these suggests that the DEEP2 field star abundances should

be scaled up by a factor of 2.8 to be representative of Stripe 82 as a whole.<sup>21</sup> A potential issue in this method of rescaling is that the stars of different magnitude need not be distributed in the same way. To test for this, we repeated the above computation for stars at fainter magnitudes, with  $22 < r < 22.5$  and found a rescaling factor of 1.4. This suggests that the larger factor is conservative, but it is also possible that the stellar density is being homogenized by galaxy contamination of the stars. The true recalibration factor for DEEP2 is probably greater than unity, but not larger than 2.8.

The statistical error on this contamination is  $\sim 10$  per cent (Poisson error); the systematic uncertainty in how it applies to a real lensing analysis, given the strong variation in stellar density across the stripe, is far larger. We therefore address the issue of real corrections for a lensing analysis in Paper II.

## 7 DISCUSSION

We have constructed deep, co-added imaging of the SDSS equatorial stripe. The procedure is designed to enable the construction of a catalogue suitable for weak lensing measurements by suppressing the effects of PSF anisotropy on the measured galaxy shapes below the level of statistical error achievable with a cosmic shear survey on this Stripe.

The galaxy density of  $\sim 2 \text{ arcmin}^{-2}$  is relatively low for a cosmic shear survey. However, it makes sense given our depth limits and large PSF of the SDSS, even by ground-based standards. As a simple point of comparison, the CMC (Jouvel et al. 2009) is commonly used to forecast galaxy yields for dark energy investigations. The effective radius of the co-added PSF is 0.67 arcsec; for Gaussians one would then expect that our cut on  $R_2 > 0.333$  should correspond to a cut on effective radius of  $r_{\text{eff}} > 0.67\sqrt{0.333/(1-0.333)} = 0.47 \text{ arcsec}$ . Using the 2011 August 15 update of the CMC, and imposing this cut as well as  $r < 23.5$  and  $i < 22.5$  (observer frame at  $A_r = 0.141$ ), we forecast a galaxy density of  $n = 2.7 \text{ arcmin}^{-2}$  and a mean source redshift  $\langle z \rangle = 0.51$ , before any small-scale masking due to e.g. bright stars and bad columns. Therefore, the final galaxy yield is broadly consistent with the tools being used to design next-generation surveys.

This procedure is successful if and only if it renders the PSF shape distortions sufficiently small that they are negligible compared to the statistical errors expected for a cosmic shear signal in this survey. To estimate the amplitude and scale dependence of the residual PSF systematics, we have measured the star–galaxy and star–star ellipticity correlation functions in our catalogue. We now fit a power law of the form:

$$\xi_{\text{sg}} = A\theta^{-p} \quad (39)$$

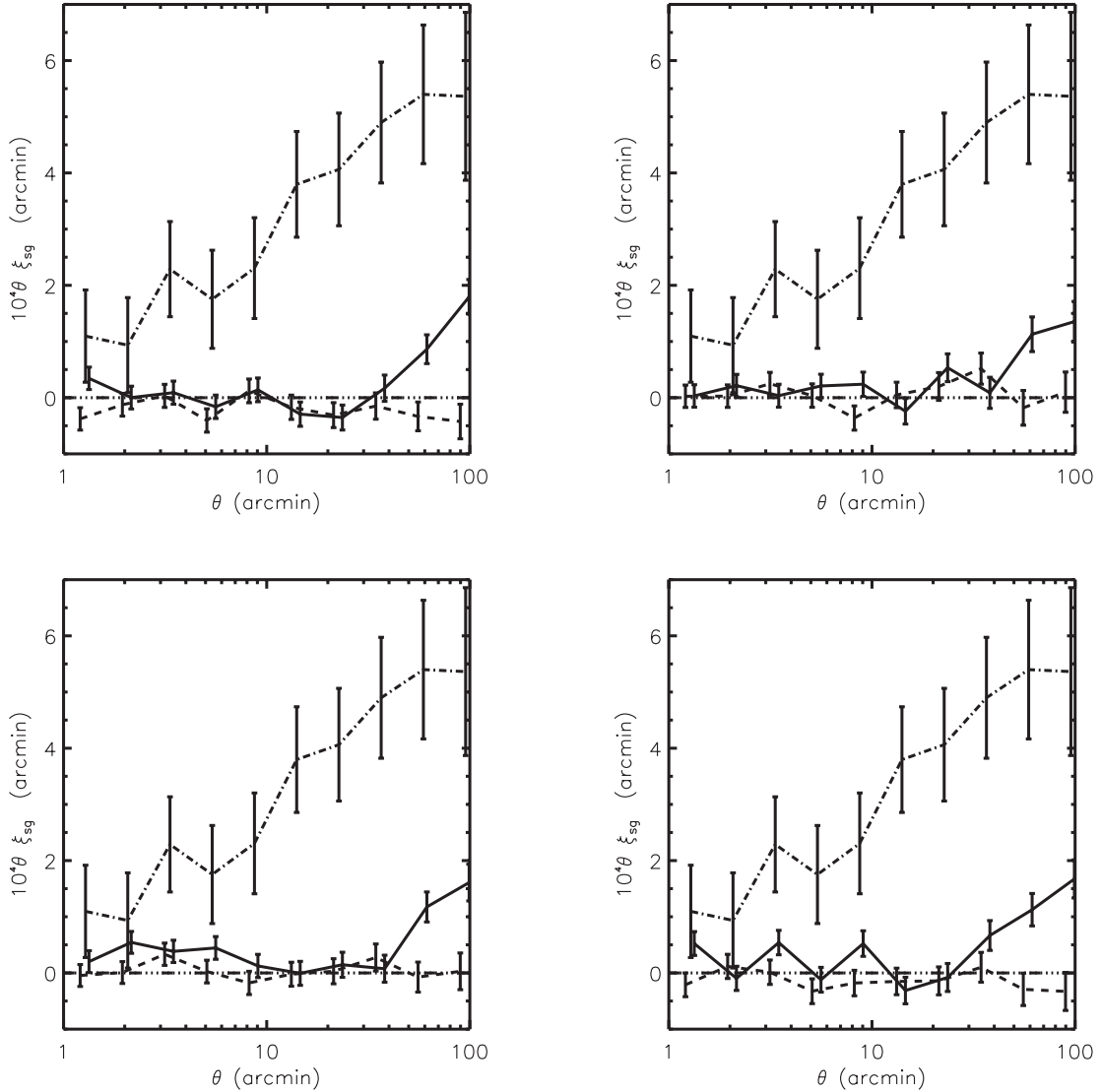
to the average of the four measured star–galaxy cross-correlations, using the Poisson errors output by the correlation function code. The best-fitting power law and average star–galaxy correlations are shown in Fig. 13, with  $(A, p) = (1.4 \times 10^{-5}, 0.85)$ .

We compare the ratio of this best-fitting power law to the shot-noise errors expected for a galaxy shape autocorrelation function for this survey. To estimate the shot noise, we follow Schneider et al. (2002) to calculate the statistical errors expected due to shot noise for a  $168$  square degree lensing survey with an effective source

<sup>20</sup> <http://deep.berkeley.edu/DR1/>

<sup>21</sup> About half of these stars are in the  $310^\circ < \text{RA} < 320^\circ$  range.





**Figure 12.** The cross-correlation of UKIDSS-selected star shapes with galaxy shapes, for the following pairs of bands: (*i*, *i*) in the upper-left, (*r*, *r*) in the upper-right, (*r*, *i*) in the bottom-left and (*i*, *r*) in the bottom-right panels. All results are shown as  $10^4 \theta \xi$ . The  $\langle e_1 e_1 \rangle$  correlation is denoted by the solid line, while the  $\langle e_2 e_2 \rangle$  correlation is shown by the dashed line. The dot-dashed line shows the expected cosmic shear  $\langle e_+ e_+ \rangle$  shape-shape correlation for a survey of this depth and size, with shot-noise errors. The triple dot-dashed line shows the ideal value of zero for the star-galaxy correlations.

surface density of 2 galaxies per arcmin<sup>2</sup>:

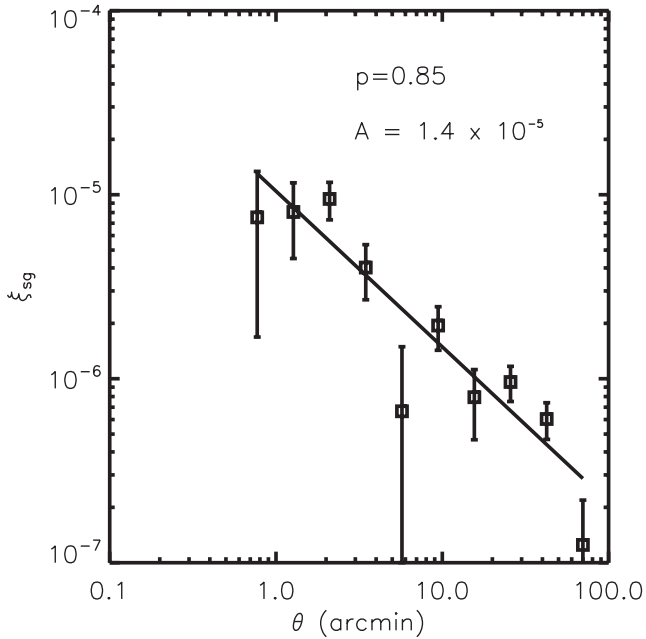
$$\text{Var}(\xi) = (3.979 \times 10^{-9}) \left( \frac{\sigma_e}{0.3} \right)^4 \left( \frac{\text{Area}}{1 \text{ deg}^2} \right)^{-1} \times \left( \frac{n_{\text{eff}}}{30 \text{ arcmin}^{-2}} \right)^{-2} \left( \frac{\theta}{1 \text{ arcmin}} \right)^{-2}. \quad (40)$$

The ratio of the systematics amplitude to the shot noise is shown as a function of scale in Fig. 14. From this, we can see that PSF systematics for these data should be, on average, 50 per cent of the size of the statistical error budget for a cosmic shear measurement with this catalogue; on degree scales, this becomes comparable to the shot-noise errors. As discussed above, this is an upper limit for three reasons: (1) imperfect star-galaxy separation at the level of several to 10 per cent can produce a star-galaxy correlation signal in the absence of uncorrected PSF effect; (2) the response of a galaxy shape to a PSF anisotropy is typically less than unity; and (3) the

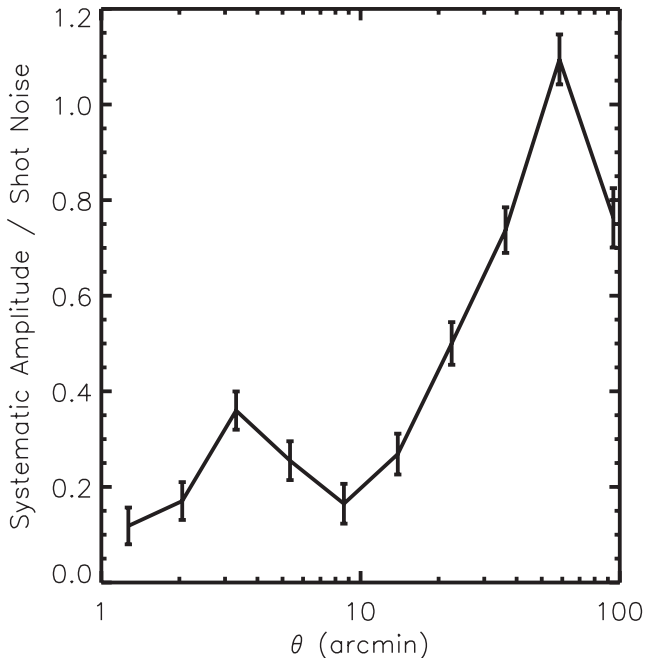
Poisson error estimate will underestimate the true variance of the signal on larger scales, where cosmic variance becomes important.

In addition, masks defined as sets of pixels can introduce a shape selection bias. We tested the effects of masking on the spurious shear statistics during the catalogue-making step by applying a strict cut to eliminate those regions of the co-add imaging with fewer than seven contributing single-epoch images. Introducing this cut actually *increased* the spurious shear amplitude; the star-galaxy correlations in the presence of this more aggressive masking step reach an amplitude of  $10^{-5}$  at degree scales.

In Appendix A, we derive the approximate expected amplitude of the spurious correlations due to pixel masking, and confirm that these should be consistent with the observed levels for this survey. The relative contributions of mask selection and PSF anisotropy biases can be ascertained empirically from the relative amplitudes of the star-star and star-galaxy correlation functions. A PSF anisotropy will produce a similar signal in both metrics. The stellar shape dispersion and typical stellar size are much smaller than



**Figure 13.** Average of the  $ri$ ,  $ir$ ,  $rr$  and  $ii$  star-galaxy cross-correlation functions, and the best-fitting power law from equation (39).



**Figure 14.** Ratio of the best-fitting star-galaxy cross-correlation power law to the expected shot-noise errors for a cosmic shear measurement using the catalogues described here. As the star-galaxy amplitude is only poorly constrained, this should be taken as a rough indication of the level of significance of the systematics.

those of the galaxies, so a selection bias will produce a much larger systematics signal in the star-galaxy correlation function than in the star-star correlation functions. This is indeed the case, as shown in Fig. 15 – substantial evidence that mask selection bias will be a significant fraction of the systematic error budget. Excluding objects near the boundaries of masked regions on the basis of their centroid positions could remove this effect; however, as Fig. 14 shows, the

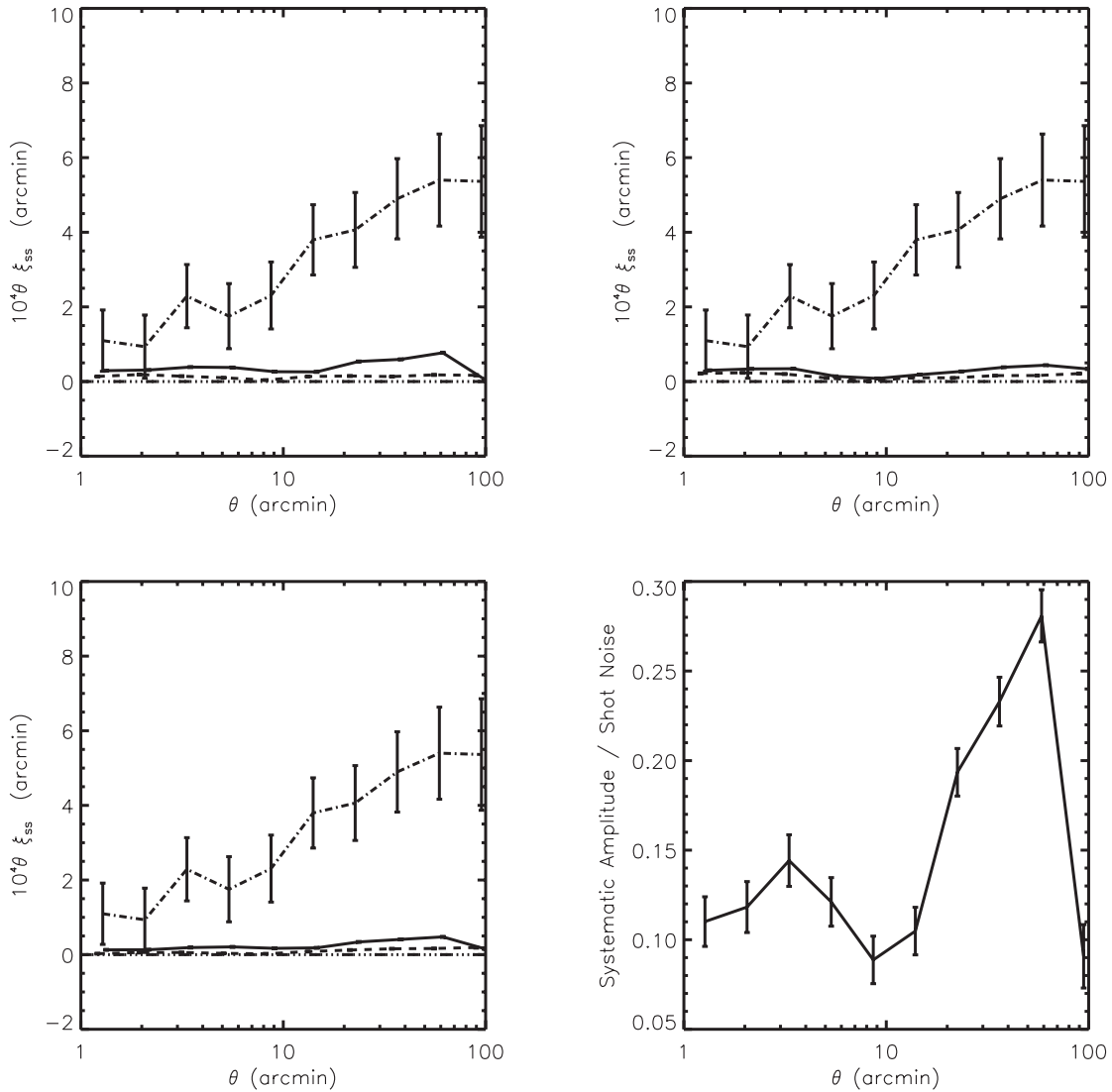
statistical errors should dominate for this catalogue, so reducing the catalogue further at this stage would not improve the quality of a final cosmic shear measurement. We also considered trying to simulate and subtract the masking bias. Ultimately, however, we settled on a more empirical approach: as described in detail in Paper II, the galaxy shear autocorrelation function used in the final analysis determines and subtracts the mean  $e_1$  in each declination bin.

These results suggest that a cosmic shear analysis that is statistics limited is possible with these data. We have shown that the effects of the PSF are small compared to the statistical errors. The masking-induced selection bias is larger, but still on average significantly smaller than the expected statistical errors. A full analysis involving the source redshift distribution, shear calibration and the cosmological implications of the two-point statistics of these data will follow in Paper II.

The systematics floor for the rounding kernel method we have employed here is set by the SDSS PSF model. Inaccuracies in this PSF model are documented both here (Fig. 8) and in other work (Reyes et al. 2012). Coherent variations in the PSF model errors in both components across the camera columns are visible with a characteristic amplitude of  $2 \times 10^{-3}$ . Aside from the very striking and atypical effect seen in the  $r$  band in camcol 2, it is likely that the shortcomings of the polynomial interpolation method employed in PHOTO play an important role here, as documented in Bergé et al. (2012) for more general simulated ground-based data. As this is close to the level of residual PSF systematics seen in our final lensing catalogue, it is very likely that an improvement in the underlying model construction would allow the rounding kernel method deployed here to achieve a greater level of systematics control.

The masking problem that we have identified is not extensively treated in the literature; to the knowledge of the authors, it has not been taken into account in existing studies. It is standard in modern photometric pipelines to define the survey mask and object rejection algorithms in terms of sets of pixels, rather than (for example) galaxy centroids, which is the ultimate source of the masking bias we see here. This effect will be important to take into account in the photometric pipeline construction in the next generation of lensing measurements. If possible, masking-related biases (and more generally, survey uniformity) should also be addressed at the observing strategy level. In this regard, the SDSS Stripe 82 technique of scanning the sky along the same guiding great circle many times, while appropriate for supernovae or transient searches, was highly non-optimal from the perspective of producing a uniform quality co-added image, since bad columns and other defects always occur at the same positions. Even dithering successive runs in the cross-scan (declination) direction by of the order of 10 arcsec would have helped this project enormously.

In Paper II, we will use the catalogue described here to measure cosmic shear. While this work was underway, we learned of a parallel effort by Lin et al. (2012). These two efforts use different methods of co-addition and different sets of cuts for the input images and galaxies; what they have in common is their use of SDSS data (not necessarily the same set of runs) and their use of the SDSS PHOTO pipeline for the initial reduction of the single-epoch data and the final reduction of the co-added data (however, they use different versions of PHOTO). Using these different methods, both groups have attempted to extract the cosmic shear signal and its cosmological interpretations. We have coordinated submission with them but have not consulted their results prior to this, so these two analysis efforts are completely independent, representing an extreme version of two independent pipelines.



**Figure 15.** The correlation functions of star shapes in the following pairs of bands:  $(r, r)$  in top-left,  $(i, i)$  in top-right and  $(r, i)$  in bottom-left panels. All results are shown as  $10^4 \theta \xi$ . The  $\langle e_1 e_1 \rangle$  correlation is denoted by the solid line, while the  $\langle e_2 e_2 \rangle$  correlation is shown by the dashed line. The dot-dashed line shows the expected cosmic shear  $\langle e_+ e_+ \rangle$  shape–shape correlation for a survey of this depth and size, with shot-noise errors. The lower-right panel shows the mean stellar cross-correlation signal as a fraction of the expected Poisson error for a cosmic shear measurement using this catalogue. The triple dot-dashed line shows the ideal value of zero for the star autocorrelations.

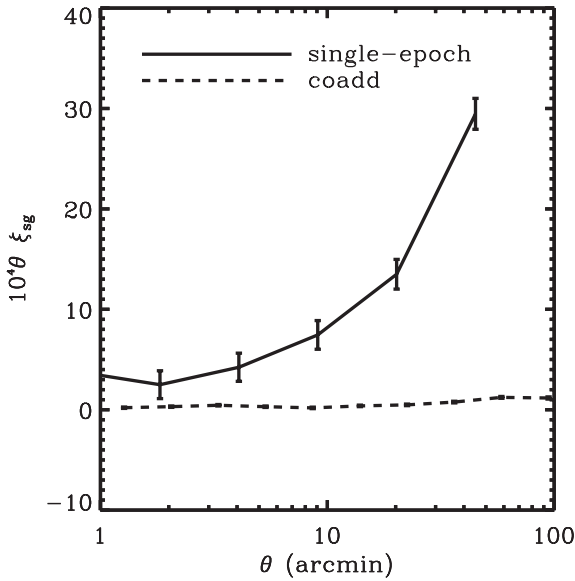
Let us consider the usable cosmological information in this measurement to be related to the total error budget, including both the systematic and statistical components. It is clear that the application of the rounding kernel to the images increases the statistical errors on the final measurement. To estimate the effects of the rounding kernel on the systematic error budget, we calculate the star–galaxy cross-correlation for single-epoch SDSS data, using the same PHOTO pipeline and shape-measurement procedure as we have for the co-adds.<sup>22</sup> The result is shown in Fig. 16. The typical amplitude of our star–galaxy cross-correlations at 40 arcmin is  $2 \times 10^{-6}$ . At the same scale, the star–galaxy correlation in the single-epoch imaging is  $7 \times 10^{-5}$ . If the PSF anisotropies in the ensemble of single-epoch images add incoherently (an overgenerous assumption, since most of the systematic PSF anisotropy appears to be aligned with the scan

direction), then we should expect the star–galaxy correlation to be smaller by roughly a factor of  $\sqrt{N}$ , where  $N = 25$  is the number of single-epoch images that passed the quality cuts on any given patch of sky, yielding an optimistic star–galaxy amplitude on this scale of  $1.4 \times 10^{-5}$ .

Secondly, the density of galaxies achieved in the final measurement ( $2.2 \text{ arcmin}^{-2}$ ) is smaller than the density predicted (again, prior to any masking) for our depth by the CMC, which was  $2.7 \text{ arcmin}^{-2}$ . The implied reduction in statistical power is about 10 per cent, for which we have gained a factor of 7 reduction in the systematic error budget. The total error in the shear measurement, including both statistical and systematic errors, is thus much smaller than that would have been the case without the co-addition and rounding kernel procedure.

We have demonstrated that a PSF homogenization method, applied to real imaging data, can produce a substantial reduction in systematic errors and a useful simplification of the shear measurement process. In SDSS this is most likely achieved because

<sup>22</sup> Note that, for the single-epoch imaging, this shape-measurement procedure applies a PSF correction using the same SDSS PSF models employed here.



**Figure 16.** The star–galaxy cross-correlation function  $\xi_{11} + \xi_{22}$  for the  $i$  band, as measured in the single-epoch SDSS lensing catalogue used for galaxy–galaxy lensing measurements as described in Reyes et al. (2012). For comparison, the star–galaxy shape cross-correlation function from the co-adds in the upper-right panel of Fig. 9 is shown as a dashed line. The same shape-measurement pipeline is used in both cases; we attribute the difference to a combination of masking effects and (more importantly) the success of the rounding kernel in suppressing those additive PSF systematics that are well described by the PSF model produced by PHOTO.

the rounding kernel method has allowed the choice of a form for the PSF that is well suited to the shape-measurement method used in the analysis.

Many aspects of the lensing analysis pipeline deployed here will not be adequate for future weak lensing surveys. The shear calibration bias is not constrainable with the SHERA simulations to the level of precision needed for LSST or *Euclid*. Our method for determining the redshift distribution of the shear catalogues is not applicable to tomographic weak lensing measurements, and may suffer from systematic errors due to spectroscopic sample selection biases that are difficult to characterize at the needed level of accuracy.

However, in future weak lensing surveys, there may be conditions under which a rounding kernel algorithm similar to that applied here will prove similarly useful.<sup>23</sup> All of these surveys will include multi-epoch data over their full footprint. In the cases where the image quality (as measured by PSF isotropy and size) distribution is sufficiently narrow (a real possibility, since the ‘shape-measurement’ bands will be acquired in the best seeing conditions), the rounding kernel method will result in little loss of information, and the freedom to choose the form of the PSF may prove useful.

## ACKNOWLEDGEMENTS

We thank Gary Bernstein, Alison Coil, Tim Eifler, Jim Gunn, Mike Jarvis, Alexie Leauthaud, Reiko Nakajima, Jeff Newman, Nikhil Padmanabhan and Barney Rowe for many useful discussions about this project. We also thank Kevin Bundy for allowing us to use preliminary versions of his UKIDSS-SDSS colour-matched catalogue.

<sup>23</sup> See Rowe, Hirata & Rhodes (2011) for an algorithm similar in spirit to that used here, but that could be applied to undersampled data.

EMH is supported by the US Department of Energy’s Office of High Energy Physics (DE-AC02-05CH11231). During the period of work on this paper, CH was supported by the US Department of Energy’s Office of High Energy Physics (DE-FG03-02-ER40701 and DE-SC0006624), the US National Science Foundation (AST-0807337), the Alfred P. Sloan Foundation and the David & Lucile Packard Foundation. RM was supported for part of the duration of this project by NASA through Hubble Fellowship grant #HST-HF-01199.02-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. US is supported by the DOE, the Swiss National Foundation under contract 200021-116696/1 and WCU grant R32-10130.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The SDSS website is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory and the University of Washington.

## REFERENCES

- Abazajian K. et al., 2003, *AJ*, 126, 2081
- Abazajian K. et al., 2004, *AJ*, 128, 502
- Abazajian K. et al., 2005, *AJ*, 129, 1755
- Abazajian K. et al., 2009, *ApJS*, 182, 543
- Adelman-McCarthy J. et al., 2006, *ApJS*, 162, 38
- Adelman-McCarthy J. et al., 2007, *ApJS*, 172, 634
- Adelman-McCarthy J. et al., 2008, *ApJS*, 175, 297
- Aihara H. et al., 2011, *ApJS*, 193, 29
- Albrecht A. et al., 2006, preprint ([arXiv:astro-ph/0609591](https://arxiv.org/abs/astro-ph/0609591))
- Bacon D., Refregier A., Ellis R., 2000, *MNRAS*, 318, 625
- Bacon D. J., Massey R. J., Refregier A. R., Ellis R. S., 2003, *MNRAS*, 344, 673
- Baldry I. K. et al., 2010, *MNRAS*, 404, 86
- Benjamin J. et al., 2007, *MNRAS*, 381, 702
- Bergé J., Price S., Amara A., Rhodes J., 2012, *MNRAS*, 419, 2356
- Bernstein G., Jarvis M., 2002, *AJ*, 123, 583
- Brown M., Taylor A., Bacon D., M., Dye S., Meisenheimer K., Wolf C., 2003, *MNRAS*, 341, 100
- Bundy K., Hogg D. W., Higgs T. D., Nichol R. C., Yasuda N., Masters K. L., Lang D., Wake D. A., 2012, *AJ*, 144, 188
- Casali M. et al., 2007, *A&A*, 467, 777
- Coil A. L., Newman J. A., Kaiser N., Davis M., Ma C.-P., Kocevski D. D., Koo D. C., 2004, *ApJ*, 617, 765
- Conroy C., Wechsler R. H., 2009, *ApJ*, 696, 620
- Frieman J. et al., 2008, *AJ*, 135, 338
- Fu L. et al., 2008, *A&A*, 479, 9



- Fukugita M., Ichikawa T., Gunn J., Doi M., Shimasaku K., Schneider D., 1996, *AJ*, 111, 1748
- Gunn J. et al., 1998, *AJ*, 116, 3040
- Gunn J. et al., 2006, *AJ*, 131, 2332
- Hambly N. C. et al., 2008, *MNRAS*, 384, 637
- Hewett P. C., Warren S. J., Leggett S. K., Hodgkin S. T., 2006, *MNRAS*, 367, 454
- Heymans C. et al., 2006, *MNRAS*, 371, L60
- Hirata C., Seljak U., 2003, *MNRAS*, 343, 459
- Hodgkin S. T., Irwin M. J., Hewett P. C., Warren S. J., 2009, *MNRAS*, 394, 675
- Hoekstra H., Yee H., Gladders M., Barrientos L., Hall P., Infante L., 2002, *ApJ*, 572, 55
- Hoekstra H., Yee H., Gladders M., 2004, *ApJ*, 606, 67
- Hoekstra H., Hsieh B. C., Yee H. K. C., Lin H., Gladders M. D., 2005, *ApJ*, 635, 73
- Hoekstra H. et al., 2006, *ApJ*, 647, 116
- Høg E. et al., 2000, *A&A*, 355, L27
- Hogg D., Finkbeiner D., Schlegel D., Gunn J., 2001, *AJ*, 122, 2129
- Huff E. M. et al., 2014, *MNRAS*, 440, 1322 (Paper II)
- Ivezić Ž. et al., 2004, *Astron. Nachr.*, 325, 583
- Jarvis M., Bernstein G., Fischer P., Smith D., Jain B., Tyson J., Wittman D., 2003, *AJ*, 125, 1014
- Jouvel S. et al., 2009, *A&A*, 504, 359
- Kaiser N., 2000, *ApJ*, 537, 555
- Kitching T. D. et al., 2012, *MNRAS*, 423, 3163
- Komatsu E. et al., 2011, *ApJS*, 192, 18
- Lawrence A. et al., 2007, *MNRAS*, 379, 1599
- Lawrence E., Heitmann K., White M., Higdon D., Wagner C., Habib S., Williams B., 2010, *ApJ*, 713, 1322
- Leauthaud A. et al., 2012, *ApJ*, 744, 159
- Lin H. et al., 2012, *ApJ*, 761, 15
- Lupton R., Gunn J., Ivezić Z., Knapp G., Kent S., 2001, in Harnden F. R., Jr, Primini F. A., Payne H. E., eds, *ASP Conf. Ser. Vol. 238, Astronomical Data Analysis Software and Systems X*. Astron. Soc. Pac., San Francisco, p. 269
- Mandelbaum R. et al., 2005, *MNRAS*, 361, 1287
- Mandelbaum R., Seljak U., Kauffmann G., Hirata C., Brinkmann J., 2006a, *MNRAS*, 368, 715
- Mandelbaum R., Seljak U., Cool R., Blanton M., Hirata C., Brinkmann J., 2006b, *MNRAS*, 372, 758
- Mandelbaum R., Li C., Kauffmann G., White S., 2009, *MNRAS*, 393, 377
- Mandelbaum R., Hirata C. M., Leauthaud A., Massey R. J., Rhodes J., 2012, *MNRAS*, 420, 1518
- Massey R. et al., 2007a, *ApJS*, 172, 239
- Massey R. et al., 2007b, *MNRAS*, 376, 13
- Massey R., Stoughton C., Leauthaud A., Rhodes J., Koekemoer A., Ellis R., Shaghoulouian E., 2010, *MNRAS*, 401, 371
- Melchior P., Böhnert A., Lombardi M., Bartelmann M., 2010, *A&A*, 510, A75
- Miyazaki S. et al., 2006, *Proc. SPIE*, 6269, 9
- Padmanabhan N. et al., 2008, *ApJ*, 674, 1217
- Pier J., Munn J., Hindsley R., Hennessy G., Kent S., Lupton R., Ivezić Ž., 2003, *AJ*, 125, 1559
- Press W., Teukolsky S., Vetterling W., Flannery B., 1992, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edn. Cambridge Univ. Press, Cambridge
- Refregier A., 2003, *MNRAS*, 338, 35
- Refregier A., Bacon D., 2003, *MNRAS*, 338, 48
- Reyes R., Mandelbaum R., Seljak U., Baldauf T., Gunn J., Lombriser L., Smith R., 2010, *Nature*, 464, 256
- Reyes R., Mandelbaum R., Gunn J. E., Nakajima R., Seljak U., Hirata C. M., 2012, *MNRAS*, 425, 2610
- Rhodes J., Refregier A., Groth E., 2001, *ApJ*, 552, L85
- Rhodes J. et al., 2007, *ApJS*, 172, 203
- Richards G. T. et al., 2002, *AJ*, 123, 2945
- Rowe B., Hirata C., Rhodes J., 2011, *ApJ*, 741, 46
- Rudd D. H., Zentner A. R., Kravtsov A. V., 2008, *ApJ*, 672, 19
- Schlegel D., Finkbeiner D., Davis M., 1998, *ApJ*, 500, 525
- Schneider P., van Waerbeke L., Kilbinger M., Mellier Y., 2002, *A&A*, 396, 1
- Schraback T. et al., 2010, *A&A*, 516, A63
- Semboloni E. et al., 2006, *A&A*, 452, 51
- Semboloni E., Schraback T., van Waerbeke L., Vafaei S., Hartlap J., Hilbert S., 2011, *MNRAS*, 410, 143
- Simha V., Weinberg D. H., Davé R., Fardal M., Katz N., Oppenheimer B. D., 2012, *MNRAS*, 423, 3458
- Smith J. et al., 2002, *AJ*, 123, 2121
- Stoughton C. et al., 2002, *AJ*, 123, 485
- Tucker D. et al., 2006, *Astron. Nachr.*, 327, 821
- Van Waerbeke L. et al., 2000, *A&A*, 358, 30
- York D. et al., 2000, *AJ*, 120, 1579
- Zacharias N. et al., 2000, *AJ*, 120, 2131
- Zhang P., Liguori M., Bean R., Dodelson S., 2007, *Phys. Rev. Lett.*, 99, 141302

## APPENDIX A: SHEAR SELECTION BIAS FROM PIXEL MASKING

In this appendix, we argue that the selection bias due to pixel masking is a quantitatively plausible explanation for the spurious shear correlations that remain after application of the PSF rounding kernel. This calculation is motivated by the observation that the spurious star–galaxy correlations increase in amplitude when more galaxies are rejected from the catalogue using the pixel mask.

Consider an elliptical galaxy isophote centred at the origin, with an ellipticity  $e$  and a major axis at some arbitrary position angle  $\theta$  with respect to the  $x$ -axis. The ellipticity  $e$  is defined in terms of the axis ratio  $q$  such that

$$e = \frac{1 - q^2}{1 + q^2}$$

$$q = \frac{b}{a}, \quad b \leq a, \quad (\text{A1})$$

where  $b$  and  $a$  are the semi-minor and semi-major axes of the ellipse, respectively.

We define the two ellipticity components  $e_+$  and  $e_-$  as

$$e_1 = e \cos(2\theta)$$

$$e_2 = e \sin(2\theta). \quad (\text{A2})$$

Naturally, the average ellipticity over all position angles  $\theta$  is zero.

Now we place a vertical barrier at position  $x = d$ ,  $d > 0$  and compute the expected value of  $e_1$  over all  $\theta$  again, this time removing the contribution from all position angles where the ellipse crosses the barrier.

Because of the  $\theta \mapsto \theta + \pi$  symmetry of ellipticity, the average value of  $e_2$  will be unchanged. When the ellipse is far enough from the barrier such that  $d > a$ , the expectation value of  $e_1$  will be zero; when the minor axis of the ellipse meets the barrier,  $d < b$ , galaxies at any position angle will be rejected from the catalogue.

The extremal point on the ellipse in the direction of the barrier is

$$x_{\max} = \sqrt{\frac{T}{2}} (1 + e_1), \quad (\text{A3})$$

where  $T = a^2 + b^2$ , and the galaxy is masked if  $x_{\max} \geq d$ . To compute the mean shape over the survey geometry, average over ellipticity weighting by the total survey area where it is possible to measure each ellipticity:

$$\langle e_1 \rangle = \frac{\int_{-1}^1 de_1 A(e_1) e_1}{\int_{-1}^1 de_1 A(e_1)}, \quad (\text{A4})$$

and  $A(e_1)$ , the total survey area where it is possible to measure an ellipticity  $e_1$ , is

$$A(e_1) = (A_{\text{tot}} - x_{\text{max}} P), \quad (\text{A5})$$

where  $A_{\text{tot}}$  is the total unmasked survey area and  $P$  is the total length of the mask perimeter. For  $|e| \ll 1$ , the mean  $e_1$  evaluates to

$$\langle e_1 \rangle \simeq -\frac{P}{A_{\text{tot}}} \sqrt{\frac{T}{2}} \left\langle \frac{e_1^2}{2} \right\rangle. \quad (\text{A6})$$

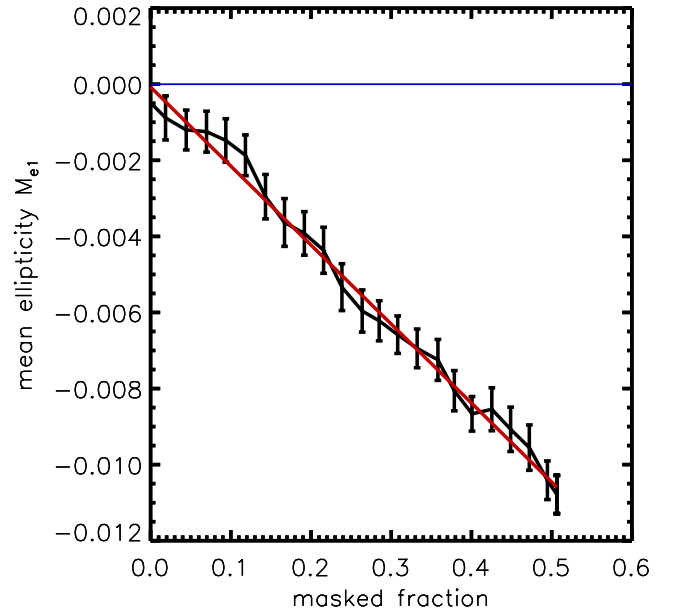
The prefactor  $\frac{P}{A_{\text{tot}}} \sqrt{\frac{T}{2}}$  is the fraction of the total survey area that lies nearer than the characteristic radius of a galaxy to a mask boundary.

For the SDSS Stripe 82 data presented here, the variance in ellipticity per component  $\langle e_1^2 \rangle$  is 0.1 and the characteristic limiting galaxy isophotal radius<sup>24</sup> is twice the median half-light radius, or 2.5 arcsec. Roughly 1 per cent of the survey area lies within this distance of a mask boundary. The shape-measurement procedure described in Section 4.9 will amplify the masking selection bias by a factor of  $R_2^{-1}$ , a characteristic value of which is  $R_2^{-1} = 3$ . This yields a mean masking selection bias of the order of  $10^{-3}$ .

To demonstrate the existence of this masking bias, we simulate it using the SHERA-based simulation code described in Paper II. COSMOS galaxies, convolved to our model PSF, are added to noisy images with the noise amplitude adjusted to match that typical of the  $r$ -band co-added images. We designate 25 equally spaced ‘bad columns’, each a single pixel wide, running parallel to the long axis in each simulated image.

We vary the masked fraction along these bad columns by only eliminating galaxies with centroid positions less than or equal to a chosen row number. Increasing the row number increases the masked fraction. The measured ellipticities here are the  $M_{\text{E1}}$  and  $M_{\text{E2}}$  moments produced by PHOTO; the PSF dilution correction has not been applied, so the effect on the dilution-corrected shapes is larger by approximately a factor of  $R_2^{-1}$ . The results, shown in Fig. A1, are broadly consistent with the analytic estimate above. With a masked fraction (along the mask perimeter) of 0.01, the best-fitting trend shown in the figure and a typical  $R_2$  of 0.333, the simulations predict a mean masking effect shape bias of 0.0006. This is large enough to dominate that part of our systematic error budget which is not accounted for by the other sources of systematic error described above, though it is still much smaller than the statistical errors.

It is important to note that, while a square or round masking pattern will generally eliminate a bias in the *average* catalogue ellipticity, this selection bias produces a coherent shape pattern along



**Figure A1.** Effects of mask boundaries on the mean ellipticity in simulated images. The horizontal axis shows that fraction of the galaxies that are detected, but are excised from the catalogue due to overlap with arbitrarily chosen masked columns a single pixel in width. The black line with errors shows the mean ellipticity, varying as the fraction of masked galaxies is changed. Bootstrap errors are shown, with one quarter of the simulated galaxy sample selected (without replacement) for each realization. Solid red line shows the best linear fit to this trend; the slope is  $-0.02076$ , and the intercept is not significantly different from zero. Note that the effect on the inferred ellipticities will be stronger by a factor of  $R_2^{-1} \simeq 3$ .

the boundary of a masked region of any size or shape. The exact effect on the shear statistics can be calculated by convolving the mask with the amplitude of the mask selection bias in each shape component, and calculating the two-point ellipticity correlation function (or other shear statistic of interest) of the resulting map; for this measurement, however, the bias is small enough that it does not contribute significantly to the shear correlation function, so we content ourselves with demonstrating its existence and order of magnitude.

<sup>24</sup> For a  $10\sigma$  circular Gaussian galaxy light profile, the outermost detectable isophote is a factor of 1.78 larger than the half-light radius.